Compact Textbooks in Mathematics

Martin Brokate
Götz Kersting

Measure and Integral

Birkhäuser

Compact Textbooks in Mathematics

Birkhäuser

Compact Textbooks in Mathematics

This textbook series presents concise introductions to current topics in mathematics and mainly addresses advanced undergraduates and master students. The concept is to offer small books covering subject matter equivalent to 2- or 3-hour lectures or seminars which are also suitable for self-study. The books provide students and teachers with new perspectives and novel approaches. They feature examples and exercises to illustrate key concepts and applications of the theoretical contents. The series also includes textbooks specifically speaking to the needs of students from other disciplines such as physics, computer science, engineering, life sciences, finance.

More information about this series at
http://www.springer.com/series/11225

Martin Brokate
Götz Kersting

Measure and Integral

Birkhäuser

Martin Brokate
TU München
Garching, Germany

Götz Kersting
Institut für Mathematik
Goethe-Universität Frankfurt
Frankfurt am Main, Germany

Revised and updated translation from the German language edition: Maß und Integral by Martin Brokate and Götz Kersting, © Springer Basel AG 2011.

ISSN 2296-4568 ISSN 2296-455X (electronic)
Compact Textbooks in Mathematics
ISBN 978-3-319-15364-3 ISBN 978-3-319-15365-0 (eBook)
DOI 10.1007/978-3-319-15365-0

Library of Congress Control Number: 2015941155

2010 Mathematics Subject Classification: 28-XX, 28-01

Springer Cham Heidelberg New York Dordrecht London

Cover design: deblik, Berlin

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This text is a translation of the German edition. It closely follows the original; some errors and misprints were corrected.

München and Frankfurt/Main
January 2015

Martin Brokate
Götz Kersting

Preface to the German Edition

Modern measure and integration theory is a prominent descendant of Cantor's set theory, and it played an important role for the formation of the latter. The roots of measure and integration theory thus are found in areas usually attributed to pure mathematics. Nevertheless, it has gained importance particularly for areas of mathematics strongly linked to applications—for functional analysis, partial differential equations, applied analysis and control theory, numerical mathematics, potential theory, ergodic theory, probability theory, and statistics. Measure and integration theory thus cannot be subsumed so easily under the paradigm pure *versus* applied mathematics (a paradigm which nowadays tends to become less and less persuasive anyway).

It is under this view that we have written our textbook. Indeed we have in mind readers who want to utilize the theory elsewhere and are interested in a concise exposition of the most important results. At the same time, we aim at presenting measure and integration theory as a coherent and transparent system of assertions on areas, volumes, and integrals. We think that this can be done in a compact manner so that it can be integrated into a standard bachelor's curriculum in mathematics.

From the standpoint of mathematics, the core of measure and integration theory has largely reached its final form. Nevertheless, we think that concerning its presentation, there is still room for accentuation. Our arrangement of the content does not follow the format chosen by other authors. Here are some explanations.

We do not start with the existence and uniqueness theorems for measures. We believe that such an approach better fits the needs of the students: Initially, the convergence results for integrals are important; the construction of measures—however nicely it works out following Carathéodory—may be postponed for the start. For this reason we treat these constructions only at the end of our textbook (which does not prevent a lecturer from reorganizing the material, of course). There we have opted for a presentation which directly leads to the goal, avoiding the usual discussions of set systems like algebras of sets, semi-rings, etc. At some other places, too, there are new features.

We do not intend to present the theory in all its ramifications. We concentrate on the core (as we understand it) and, beyond that, display results which provide links to other areas of mathematics. Regarding analysis, this pertains, e.g., to the smoothing of functions by convolution as well as Jacobi's transformation formula. Concerning geometric measure theory, we discuss the Hausdorff measure and dimension. For probability theory, among other things, we treat kernels and measures on infinite products following Kolmogorov. With the final two chapters, we try to exhibit some connections to functional analysis which we find useful for understanding measure and integration theory. To guide the reader, we have marked some sections with an asterisk ($*$); they may be skipped at first reading.

As a prerequisite, we assume knowledge of the contents of the first-year bachelor courses in mathematics (as they are typically given in our home country). From topology, without comment we only use elementary concepts (open, closed, compact, neighborhood, continuity) in the setting of metric spaces. Anything exceeding that, we discuss by some means or other. Historical notes are found in footnotes.

A concise text as the one we aimed at cannot substitute any comprehensive exposition. We therefore do not intend to replace established textbooks like Elstrodt's [2], much less classical texts like those of Halmos [4] or Bauer [1]. In the appendix we mention these and other introductions to the theory. From all of them, we have benefitted a lot; we take the liberty not to document this in detail, as should be permitted in a textbook. We gladly have incorporated suggestions for the text as well as corrections due to Christian Böinghoff and Henning Sulzbach. We thank Birkhäuser for the pleasant and smooth collaboration.

München and Frankfurt/Main — Martin Brokate
March 2010 — Götz Kersting

Contents

Introduction

1

To determine specific contents of area and volume as well as integrals is a very old theme in mathematics. Unsurpassed are the achievements of Archimedes, in particular his computation of the volume of the unit ball as $4\pi/3$ and of the area of the unit sphere as 4π. Starting from Euler, problems like determining the value of $\int_0^\infty \frac{\sin x}{x} dx$ (which is $\pi/2$) have kept the analysts busy.

To the end of the nineteenth century, this subject became less and less important, as there was not much left to be discovered. At that moment, measure and integration theory entered the stage. It, too, deals with contents or (as we will call it in the following) *measures* of sets, as well as with integrals of functions, but the question has changed. It no longer reads "what is the measure of this or that set?" but rather "which sets can be measured, which functions can be integrated?". To which sets one thus can assign a measure, to which functions an integral? Their specific value becomes secondary, general rules of integration come to the fore. The relation to differential calculus, which for a long period since Newton and Leibniz was in the foreground, loses its dominant role.

Such a change of perspective is not uncommon in mathematics. In our case, it arose in the context that one no longer considered integrals on their own, but rather needed them as tools in other mathematical investigations. Historically one should mention in particular the Fourier analysis of functions, the decomposition of real-valued functions into sinusoidal oscillations. Their coefficients (amplitudes) can be expressed by certain integrals—soon, one realized that for this purpose one needed properties of integration which could not be provided by the notions of integrals being available at that time.

Measure and integration theory according to Lebesgue arose by and large between the years 1900 and 1915, based on essential preliminary work of Borel[1]

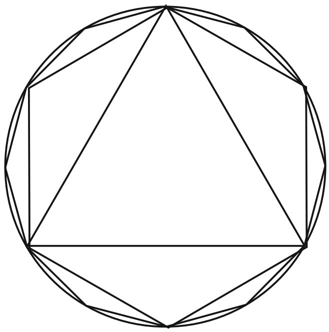
[1]Émile Borel, 1871–1956, born in Saint-Affrique, active in Paris at the École Normale Supérieure and the Sorbonne. He significantly contributed not only to the foundations of measure theory, but also to complex analysis, set theory, probability theory, and to applications of

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_1

from 1894. Right from the start, the pioneers during that time directed their attention towards the fundamental properties of measure and integral. Borel was the first to demand that measures should be not only additive, but also σ-additive. This means that not only for *finitely many* disjoint measurable sets $B_1, B_2, \ldots \subset \mathbb{R}^d$ with measures $\lambda(B_1), \lambda(B_2), \ldots$ the union $B = B_1 \cup B_2 \cup \cdots$ is measurable and has measure

$$\lambda(B) = \lambda(B_1) + \lambda(B_2) + \cdots ,$$

but that moreover this property holds for every *infinite* sequence $B_1, B_2, \ldots$ of disjoint measurable sets. Borel realized that only under this assumption a fertile mathematical theory arises. Particular cases like the circle in the figure



of course do not yield anything new. Lebesgue,[2] the founder of modern integration theory, in his fundamental treatise on integration from the year 1901 started from six properties that integrals should reasonably satisfy.

Measure and integration theory is based on set theory and cannot dispense with its ways of reasoning. Only with the aid of set theory a path was found leading to the full system of measurable subsets of $\mathbb{R}^d$ and of other spaces. Yet this approach is comparatively abstract and indirect. To realize that it is justified, for a start it is perhaps appropriate to take a look at other more descriptive approaches, even though they finally were not conclusive.

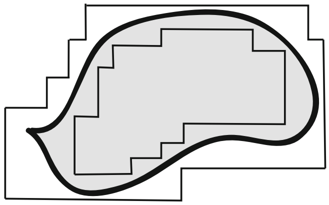
mathematics. He combined this work with a political career as member of the parliament, minister of the navy, and finally member of the Résistance.

[2] Henri Lebesgue, 1875–1941, born in Beauvais, active in Paris at the Sorbonne and the Collège de France. His foundation of integration theory is a landmark in mathematics, he could resort to preliminary work of Borel and Baire. With his methods he then obtained results on Fourier series.

Let us look at the approach due to Jordan.[3] His idea is intuitively appealing: Let $V = \bigcup_{j=1}^{k} I_j$ be a union of finitely many disjoint d-dimensional intervals $I_j \subset \mathbb{R}^d$, thus $I_j = [a_{j1}, b_{j1}) \times \cdots \times [a_{jd}, b_{jd})$ (it turns out to be useful, though not strictly necessary, to work with semi-open intervals). One obtains its measure $\lambda(V)$ by adding the products of the edge lengths of the individual intervals:

$$\lambda(V) := \sum_{j=1}^{k} (b_{j1} - a_{j1}) \cdots (b_{jd} - a_{jd}) .$$

Following Jordan, the exterior and the interior measure of a subset $B \subset \mathbb{R}^d$ result from covering resp. exhausting B by a union of intervals:



Expressed in formulas,

$$\lambda^*(B) := \inf\{\lambda(V) : V \supset B\} , \quad \lambda_*(B) := \sup\{\lambda(V) : V \subset B\} .$$

If both expressions have the same value, then the set B is called a Jordan set, and $\lambda(B) := \lambda^*(B) = \lambda_*(B)$ is called the Jordan measure of B. This definition is analogous to that of the Riemann integral of a function.

Without a doubt, this approach assigns to a Jordan set its "correct" measure. The deficiency of this approach lies elsewhere, on the structural level. Indeed, finite unions, finite intersections, and complements of Jordan sets are again Jordan sets. But it turns out that, in general, a countable union of Jordan sets is not necessarily a Jordan set. One easily sees, for example, that every set which consists of just a single element is a Jordan set of measure 0, while the set of rational numbers in [0, 1] is not a Jordan set (its inner and outer measures are 0 resp. 1). The σ-additivity is lacking.

This deficiency is fatal. All attempts to modify Jordan's definition in order to remove this deficiency have failed.

But perhaps it is not really necessary to define measurability of sets through an explicit construction. Is it maybe possible to assign a measure to *each* subset of

[3]CAMILLE JORDAN, 1838–1922, born in Lyon, active in Paris at the École Polytechnique and the Collège de France. Better known than his contributions to measure theory is his work on group theory. The Jordan normal form of matrices as well as Jordan curves demonstrate his wide mathematical interests.

$\mathbb{R}^d$ in a reasonable manner, no matter whether in a direct or an indirect fashion? Already Lebesgue posed that question. The answer is negative, as was discovered by Vitali[4] and Hausdorff.[5] Later, Hausdorff's result was extended by Banach[6] and Tarski.[7] It is somewhat perplexing and thus nowadays known as the *Banach-Tarski paradox*. These two mathematicians proved in 1924: Any two bounded subsets B and B′ of $\mathbb{R}^d$, $d \geq 3$, with nonempty interior, for example two balls of different radii, can be decomposed into an equal number of disjoint subsets $B = C_1 \cup \cdots \cup C_k$ and $B' = C'_1 \cup \cdots \cup C'_k$ such that all the parts $C_1, \ldots, C_k, C'_1, \ldots, C'_k$ are pairwise congruent, that is, they can be transformed into each other by translations, rotations and reflexions. One then is inclined to conclude that all parts have the same measure due to congruency, and therefore B and B′ have the same measure by virtue of additivity. This would be paradoxical. How can one realise such decompositions? Intuitively this is inconceivable.

The answer is the following: The theorem of Banach-Tarski is a result of set theory, and set theory (in particular, when the axiom of choice is employed) admits the formation of rather exotic subsets of $\mathbb{R}^d$ which are no longer accessible through imagination. This is the meaning of the theorem: The system of all subsets of $\mathbb{R}^d$ is so extensive that it is impossible to assign measures to every subset such that they are invariant under congruency as well as additive. Therefore, the conclusion mentioned above cannot be drawn. Thus the paradox dissolves. These results due to Vitali, Hausdorff, Banach and Tarski are significant in the history of measure theory; nowadays they rather are a special theme.

Let us record: Attempting to view measurable subsets as single items does not lead to a sound mathematical theory. Therefore, we no longer look at individual subsets, but focus instead on systems $\mathcal{B}$ of measurable subsets. Their properties are simple. Following Borel, two properties are indispensable:

$$B \in \mathcal{B} \;\Rightarrow\; B^c \in \mathcal{B} \quad \text{and} \quad B_1, B_2, \ldots \in \mathcal{B} \;\Rightarrow\; \bigcup_{n \geq 1} B_n \in \mathcal{B}$$

[4] Giuseppe Vitali, 1875–1932, born in Ravenna, active in Modena, Padova and Bologna. He provided distinguished contributions to measure theory, but also to complex analysis.

[5] Felix Hausdorff, 1868–1942, born in Breslau, active in Leipzig, Greifswald, and Bonn. Hausdorff made fundamental contributions to set theory, topology, and measure theory. His monograph on set theory had enormous influence. Under the alias *Paul Mongré* he published essayistic and literary works. Due to his Jewish origin, Hausdorff was forced to retire in 1935. To escape deportation he took his own life in 1942.

[6] Stefan Banach, 1892–1945, born in Krakow, active in Lemberg. He established modern functional analysis. The Lemberg school of mathematicians formed around him and Hugo Steinhaus.

[7] Alfred Tarski, 1902–1983, born in Warsaw, active in Warsaw and Berkeley. He is regarded as one of the most famous logicians due to, for instance, his papers on model theory. He also contributed to set theory, measure theory, algebra, and topology. Because of his Jewish origin, after the German invasion of Poland he remained in the United States.

must hold for the complement B^c of B and for finite as well as infinite sequences $B_1, B_2, \ldots$ Such systems of sets are of fundamental importance in measure theory; following Hausdorff, they are called σ-algebras. Now the task arises to exhibit a σ-algebra which is large enough and is such that σ-additivity holds when assigning a measure to its elements.

This task can be tackled in different ways. One possibility is to start from a system $\mathcal{E}$ of sets to which a measure can be assigned in an obvious manner. For this, the system of all (semi-open) intervals of $\mathbb{R}^d$ qualifies. One then enlarges $\mathcal{E}$ to the system $\mathcal{E}'$ of all countable unions of sets from $\mathcal{E}$ together with the complements of those unions. Using σ-additivity, a measure can be assigned to all elements of $\mathcal{E}'$. If $\mathcal{E}'$ is not yet a σ-algebra, one repeats this step until a σ-algebra $\mathcal{B}^d$ has emerged. This path can be (and initially has been) entered, however it turns out that uncountably many steps are required to attain the goal. This not only stresses our intuition, but moreover one has to utilize advanced methods of set theory, namely, the theory of well-ordered sets and transfinite induction. No view emerges of how a typical measurable set looks like.

Fortunately, an elementary and much simpler approach was found soon: one directly focuses on $\mathcal{B}^d$ by characterizing it as the *smallest* σ-algebra which contains $\mathcal{E}$. It is called the *Borel σ-algebra*, and its elements $B \subset \mathbb{R}^d$ are called *Borel sets*. We will see how one assigns a measure to every Borel set so that σ-additivity holds, and how an integration theory is established whose rules are transparent and easy to apply.
One has to pay a price: in order to smoothly manipulate measurable sets and integrable functions one also has to deal with sets and functions, which in no way conform to classical perceptions. Back then, leading mathematicians faced this development in a reserved or even hostile manner, Hermite,[8] for example, spoke about the "deplorable plague" of functions not possessing derivatives. Nevertheless, the ideas of Borel and Lebesgue prevailed. Their theory is one of the most important accomplishments of set theory.

As individual elements, measurable sets can hardly be controlled, one gets hold of them only through their affiliation to systems of sets. This also means that nobody can say how a "typical" Borel set looks like. In contrast, one may imagine of a typical Jordan set as the above figure suggests. Nevertheless, in the following we will no longer bring up Jordan sets, while Borel sets will remain in the focus of our considerations. In measure and integration theory one has to get used to operate with systems of sets and of functions, not with individual sets and functions.
Since its emergence, during the age of Newton and Leibniz, the integral has evolved into a fundamental tool to be employed in many areas within and outside of mathematics. Among them are the description of processes taking place in the continuum—e.g. the space-time continuum—in the corresponding areas of

[8] CHARLES HERMITE, 1822–1901, born in Dieuze, active in Paris at the École Polytechnique and at the Sorbonne. He significantly contributed to algebra and number theory, orthogonal polynomials, and elliptic functions.

(mathematical) analysis, the description of random phenomena in probability theory, as well as the description of algorithms for computer approximation and simulation of such processes in numerical mathematics and scientific computing.

In all those contexts the Lebesgue integral has turned out to be the most adequate notion of an integral. Concerning analysis and numerical mathematics, the main reason is that the functions whose p-th power possesses a Lebesgue integral form a complete space (that is, every Cauchy sequence converges) with respect to the integral norm. In the case $p = 2$ the integral moreover yields a scalar product, and we obtain a Hilbert space. These spaces, called L_p spaces, and their descendants—for example, the Sobolev spaces—provide the predominant mathematical framework for problems in the continuum.

While Lebesgue integration theory does not concern itself with the computation of specific integrals, some of its results nevertheless assist this purpose. The results pertaining to the interchange of integrals and limits (on monotone and dominated convergence) have manifold applications, for example they clarify under which circumstances derivatives and integrals can be interchanged. Analogously, this is true for the theorems of Fubini[9] and Tonelli[10] concerning interchanging the order of integration for multiple integrals. Some specific important integrals will be dealt with in the text.

---

[9] GUIDO FUBINI, 1879–1943, born in Venice, active in Catania, Turin, and Princeton. He worked on real analysis, differential geometry, and complex analysis. 1939 he emigrated to the USA after he had lost his chair in Turin in the course of the antisemitic politics under Mussolini.

[10] LEONIDA TONELLI, 1885–1946, born in Gallipoli near Lecce, active in Cagliari, Parma, Bologna, and Pisa. He worked in many areas of analysis and is known mainly for his contributions to the calculus of variations.

2 Measurability

In this chapter we introduce measurable sets and measurable functions. As explained in the introduction, the objects we operate with are mainly systems of sets, and not individual sets. In doing so, there will arise finite as well as infinite sequences of sets. In both cases and, regardless of their length, we denote such sequences as $A_1, A_2, \ldots$, their union as $\bigcup_{n\geq 1} A_n$, and so on.

Definition

A system $\mathcal{A}$ of subsets of a nonempty set S with the properties

(i) $S \in \mathcal{A}$,
(ii) $A \in \mathcal{A} \quad \Rightarrow \quad A^c := S \setminus A \in \mathcal{A}$,
(iii) $A_1, A_2, \ldots \in \mathcal{A} \quad \Rightarrow \quad \bigcup_{n\geq 1} A_n \in \mathcal{A}$,

is called a σ-*algebra* or a σ-*field* in S. The pair $(S, \mathcal{A})$ is called a *measurable space*. The elements of $\mathcal{A}$ are termed *measurable subsets* of S.

As a consequence,

(iv) $\emptyset = S^c \in \mathcal{A}$,
(v) $A_1, A_2, \ldots \in \mathcal{A} \quad \Rightarrow \quad \bigcap_{n\geq 1} A_n = (\bigcup_{n\geq 1} A_n^c)^c \in \mathcal{A}$,
(vi) $A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \setminus A_2 := A_1 \cap A_2^c \in \mathcal{A}$,
(vii) $A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \Delta A_2 := (A_1 \cup A_2) \setminus (A_1 \cap A_2) \in \mathcal{A}$.

Definition

Let $(S, \mathcal{A})$, $(S', \mathcal{A}')$ be measurable spaces. A mapping $\varphi : S \to S'$ is called *measurable*, more precisely $\mathcal{A}$-$\mathcal{A}'$-*measurable*, if preimages of measurable sets are themselves measurable, that is, if

$$\varphi^{-1}(A') \in \mathcal{A} \quad \text{for all } A' \in \mathcal{A}'.$$

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_2

When dealing with a measurable space based on a set S, the choice of the σ-algebra $\mathcal{A}$ will usually be obvious, and thus it will be clear which subsets of S are measurable. Therefore, in the following we will not always specify $\mathcal{A}$ explicitly.

Example (Trace σ-Algebra)

Given a measurable subset S_1 in a measurable space $(S, \mathcal{A})$, the system $\mathcal{A}_1 := \{A \subset S_1 : A \in \mathcal{A}\}$ becomes a σ-algebra on S_1. It is called the *trace σ-algebra* of $\mathcal{A}$ on S_1, or alternatively the *induced σ-algebra* on S_1. A mapping $\varphi : S \to S'$ is $\mathcal{A}$-$\mathcal{A}'$-measurable if and only if the restrictions of φ on S_1 and on $S_2 := S_1^c$ are measurable with respect to the trace σ-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, thanks to the formula

$$\varphi^{-1}(A') = (\varphi^{-1}(A') \cap S_1) \cup (\varphi^{-1}(A') \cap S_2) .$$

Proposition 2.1 (Composition of measurable mappings). *Let* $(S, \mathcal{A})$*,* $(S', \mathcal{A}')$ *and* $(S'', \mathcal{A}'')$ *be measurable spaces,* $\varphi : S \to S'$ *be a* $\mathcal{A}$*-*$\mathcal{A}'$*-measurable mapping, and* $\psi : S' \to S''$ *be a* $\mathcal{A}'$*-*$\mathcal{A}''$*-measurable mapping. Then* $\psi \circ \varphi : S \to S''$ *is a* $\mathcal{A}$*-*$\mathcal{A}''$*-measurable mapping.*

Proof. For any measurable subset A'' of S'', the set $A' := \psi^{-1}(A'')$ is measurable in S' by assumption, and therefore $(\psi \circ \varphi)^{-1}(A'') = \varphi^{-1}(A')$ is measurable in S. □

Generators of σ-Algebras, Borel σ-Algebras

When the set S is countable, the usual choice of the σ-algebra is the power set, the set of all subsets of S. For uncountable sets S, however, this approach has turned out to be unsuitable. Instead, in that case one specifies a σ-algebra by a generator.

Definition

A system $\mathcal{E}$ of subsets of S is called a *generator* of the σ-algebra $\mathcal{A}$ in S, if $\mathcal{A}$ is the smallest σ-algebra in S which contains $\mathcal{E}$ (that is, if for every σ-algebra $\tilde{\mathcal{A}}$ on S with $\tilde{\mathcal{A}} \supset \mathcal{E}$ we also have $\tilde{\mathcal{A}} \supset \mathcal{A}$). $\mathcal{A}$ is called the *σ-algebra generated by* $\mathcal{E}$ and is denoted by $\mathcal{A} = \sigma(\mathcal{E})$.

Every system of subsets generates a σ-algebra.

Proposition 2.2 (Generated σ-algebras). *For every system* $\mathcal{E}$ *of subsets in* S *there is a smallest σ-algebra which contains* $\mathcal{E}$*. We obtain it as the intersection of all σ-algebras containing* $\mathcal{E}$*:*

$$\sigma(\mathcal{E}) = \{A \subset S \;:\; A \in \tilde{\mathcal{A}} \text{ for every σ-algebra } \tilde{\mathcal{A}} \text{ in } S \text{ with } \tilde{\mathcal{A}} \supset \mathcal{E}\} .$$

Proof. The system of all σ-algebras containing $\mathcal{E}$ is nonempty, since the system of *all* subsets of S belongs to it. The intersection $\mathcal{A}$ of all those σ-algebras is itself a σ-algebra. Indeed, $A \in \mathcal{A}$ means that $A \in \tilde{\mathcal{A}}$ for all σ-algebras $\tilde{\mathcal{A}} \supset \mathcal{E}$. It follows that $A^c \in \tilde{\mathcal{A}}$ for all $\tilde{\mathcal{A}} \supset \mathcal{E}$ and therefore $A^c \in \mathcal{A}$. The other properties of a σ-algebra are derived analogously. Moreover, we have $\mathcal{A} \supset \mathcal{E}$ as well as $\mathcal{A} \subset \tilde{\mathcal{A}}$ for every σ-algebra $\tilde{\mathcal{A}} \supset \mathcal{E}$. The proposition is proved. □

When working with generated σ-algebras the following statements are used routinely.

Proposition 2.3 (Equality of σ-algebras). *Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be generators of the σ-algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ in* S. *Then $\mathcal{A}_1 = \mathcal{A}_2$ holds if $\mathcal{E}_1 \subset \mathcal{A}_2$ and $\mathcal{E}_2 \subset \mathcal{A}_1$.*

Proof. From $\mathcal{E}_1 \subset \mathcal{A}_2$ we conclude that $\mathcal{A}_1 \subset \mathcal{A}_2$, and vice versa. □

Proposition 2.4 (Measurability criterion). *Let* $(S, \mathcal{A})$, $(S', \mathcal{A}')$ *be measurable spaces, and let $\mathcal{E}'$ generate $\mathcal{A}'$. Then $\varphi : S \to S'$ is an $\mathcal{A}$-$\mathcal{A}'$-measurable mapping, if*

$$\varphi^{-1}(A') \in \mathcal{A} \quad \textit{for all } A' \in \mathcal{E}'.$$

Proof. $\tilde{\mathcal{A}} := \{A' \in \mathcal{A}' : \varphi^{-1}(A') \in \mathcal{A}\}$ is a σ-algebra, as a brief computation shows. By assumption, $\mathcal{E}' \subset \tilde{\mathcal{A}} \subset \mathcal{A}'$. Since $\mathcal{A}'$ is the smallest σ-algebra containing $\mathcal{E}'$, we conclude that $\tilde{\mathcal{A}} = \mathcal{A}'$, and the assertion follows. □

Rather frequently one considers the σ-algebra generated by the open subsets in a Euclidean space or, more generally, in a metric space.

Definition

Let (S, d) be a metric space with metric d and let $\mathcal{O}$ be the system of its open subsets. Its *Borel-σ-algebra* $\mathcal{B} := \sigma(\mathcal{O})$ is defined as the σ-algebra generated by the open subsets of S. Its elements are called *Borel sets*. A mapping between two metric spaces is called *Borel measurable*, if it is measurable w.r.t. the Borel σ-algebras.

In a topological space, too, the σ-algebra generated by the open sets is called the Borel σ-algebra. We restrict our treatment to metric spaces, where the circumstances remain clear.

We now have at our disposal a method for constructing measurable sets which is highly indirect. In general it does not give us any indication which subsets of S actually belong to $\sigma(\mathcal{E})$ resp. $\sigma(\mathcal{O})$. In contrast to, e.g., the open sets in a metric space, they cannot be characterized "individually". However, this does not create any serious problems; one just works with systems of sets instead of individual sets.

Example

1. By virtue of Proposition 2.3, the Borel σ-algebra is also generated by the system of all closed subsets (the complements of the open sets).
2. Every continuous mapping between two metric spaces is Borel measurable. This follows from Proposition 2.4 because, for continuous mappings, the preimages of open sets are again open and hence Borel measurable.
3. We denote the Borel σ-algebra of the Euclidean space $\mathbb{R}^d$ by $\mathcal{B}^d$. It is generated, too, by the system of all d-dimensional open intervals of the form

$$(-\infty, b) := (-\infty, b_1) \times \cdots \times (-\infty, b_d) , \quad b = (b_1, \ldots, b_d) \in \mathbb{R}^d .$$

Indeed, from those intervals we may obtain every finite half-open interval $[a, b) = [a_1, b_1) \times \cdots \times [a_d, b_d)$ according to

$$[a, b) = (-\infty, b) \setminus \bigcup_{i=1}^{d} (-\infty, c_i)$$

with $c_i := (b_1, \ldots, b_{i-1}, a_i, b_{i+1}, \ldots, b_d)$, and furthermore every open set O as a countable union of half-open intervals according to

$$O = \bigcup \left\{[a, b) \ : \ [a, b) \subset O, \ a, b \in \mathbb{Q}^d\right\} ;$$

note that, because the rational numbers are dense in $\mathbb{R}$, for every open set O and every $x \in O$ there is an interval $[a, b)$ with $x \in [a, b) \subset O$ and $a, b \in \mathbb{Q}^d$. Therefore, the finite half-open intervals $[a, b)$ generate the Borel σ-algebra, too. In the same manner $\mathcal{B}^d$ is also generated by all finite open or by all finite closed intervals, and moreover by all intervals $(-\infty, b]$, $b \in \mathbb{R}^d$.
4. As a consequence, every monotone mapping $\varphi : \mathbb{R} \to \mathbb{R}$ is Borel measurable, since the preimage of an interval under φ is again an interval, and hence a Borel set.
5. Let $\varphi_1, \varphi_2, \ldots$ be an infinite sequence of measurable mappings from a measurable space S with σ-algebra $\mathcal{A}$ to a metric space S' with metric d and Borel σ-algebra $\mathcal{B}$. We assume that the sequence converges pointwise to a mapping $\varphi : S \to S'$, thus $d(\varphi_n(x), \varphi(x)) \to 0$ holds for all $x \in S$. Then φ is measurable. Indeed, let $B \subset S'$, $\varepsilon > 0$, and let $U_\varepsilon(B) := \{y \in S' : d(y, z) < \varepsilon \text{ for a } z \in B\}$ be the "open ε-neighbourhood" of B. If B is closed, then for every seqence $\varepsilon_1 \geq \varepsilon_2 \geq \cdots > 0$ converging to 0 we have

$$\varphi^{-1}(B) = \bigcap_{k=1}^{\infty} \{x \in S : \varphi_n(x) \in U_{\varepsilon_k}(B) \text{ except for finitely many } n\}$$
$$= \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \{x \in S : \varphi_n(x) \in U_{\varepsilon_k}(B) \text{ for all } n \geq m\}$$
$$= \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \varphi_n^{-1}\big(U_{\varepsilon_k}(B)\big) \in \mathcal{A} ,$$

and the assertion follows from Proposition 2.4. This convergence property is a feature which distinguishes the class of measurable functions from other classes of functions (like e.g., the continuous functions), compare Exercise 7.4.

We may also generate σ-algebras from mappings.

Definition

Let $(S_i, \mathcal{A}_i)$, $i \in I$, be measurable spaces and $\psi_i : S' \to S_i$, $i \in I$, mappings. The smallest σ-algebra $\mathcal{A}'$ in S' such that all ψ_i are $\mathcal{A}'$-$\mathcal{A}_i$-measurable mappings is called the *σ-algebra generated by the* (ψ_i). We denote it by $\mathcal{A}' = \sigma(\psi_i, i \in I)$.

The σ-algebra $\sigma(\psi_i, i \in I)$ is generated by $\mathcal{E}' = \bigcup_{i \in I} \{\psi_i^{-1}(A_i) : A_i \in \mathcal{A}_i\}$.

Example

The Borel σ-algebra $\mathcal{B}$ in a metric space S with metric d coincides with the σ-algebra $\mathcal{B}'$ generated by all continuous functions $\psi : S \to \mathbb{R}$. On the one hand, continuous functions are Borel measurable, therefore $\mathcal{B}' \subset \mathcal{B}$. On the other hand, for all sets $B \subset S$ the function $x \mapsto \psi_B(x) := \inf\{d(x,z) : z \in B\}$ (the "distance" between x and B) is continuous from S to $\mathbb{R}$, because $|\psi_B(x) - \psi_B(y)| \leq d(x,y)$. If B is closed we have in addition that $x \in B \Leftrightarrow \psi_B(x) = 0$, thus $B = \psi_B^{-1}(\{0\})$. Therefore, $\mathcal{B}'$ includes all closed sets, and by Proposition 2.3 we conclude that $\mathcal{B} \subset \mathcal{B}'$.

The following statement corresponds to the measurability criterion.

Proposition 2.5. *Let* $(S, \mathcal{A})$, $(S', \mathcal{A}')$ *and* $(S_i, \mathcal{A}_i)$, $i \in I$, *be measurable spaces, and let* $\mathcal{A}'$ *be generated by the mappings* $\psi_i : S' \to S_i$, $i \in I$. *A mapping* $\varphi : S \to S'$ *is* $\mathcal{A}$-$\mathcal{A}'$*-measurable if and only if* $\psi_i \circ \varphi$ *is* $\mathcal{A}$-$\mathcal{A}_i$*-measurable for all* i.

Proof. The "only if"-part follows because the composition of measurable mappings is again measurable. For the converse, let $\psi_i \circ \varphi$ be measurable for all i, thus $(\psi_i \circ \varphi)^{-1}(A_i) \in \mathcal{A}$ for all $A_i \in \mathcal{A}_i$. This means that $\varphi^{-1}(A') \in \mathcal{A}$ for all $A' = \psi_i^{-1}(A_i)$ with $A_i \in \mathcal{A}_i$. Those sets A' generate the σ-algebra $\mathcal{A}'$. The measurability of φ now follows from the measurability criterion. □

Product Spaces

We now apply our construction method for σ-algebras to finite or countably infinite Cartesian products

$$S_\times = \prod_{n\geq 1} S_n = S_1 \times S_2 \times \cdots$$

Let $\mathcal{A}_1, \mathcal{A}_2, \ldots$ be σ-algebras on $S_1, S_2, \ldots$ We call a subset of $S_\times$ of the form

$$A_1 \times A_2 \times \cdots \quad \text{with} \quad A_n \in \mathcal{A}_n$$

a *measurable rectangle*.

Definition

The σ-algebra $\mathcal{A}_\otimes$ in $S_\times$ generated by all measurable rectangles is called the *product σ-algebra* of $\mathcal{A}_1, \mathcal{A}_2, \ldots$. We call $(S_\times, \mathcal{A}_\otimes)$ the *product space* of $(S_n, \mathcal{A}_n)$ and write

$$\mathcal{A}_\otimes = \bigotimes_{n\geq 1} \mathcal{A}_n = \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \cdots .$$

If, in particular, $S_1 = S_2 = \cdots = S$ und $\mathcal{A}_1 = \mathcal{A}_2 = \cdots = \mathcal{A}$, we write

$$S^d \text{ and } \mathcal{A}^d \tag{2.1}$$

instead of $S_\times$ and $\mathcal{A}_\otimes$. Here d denotes the length of the sequence $S_1, S_2, \ldots$ The case $d = \infty$ is included; S^∞ is just the set of infinite sequences in S.

Alternatively, we may describe the product σ-algebra by the *projection mappings* $\pi_i : S_\times \to S_i$, $i \geq 1$, given by

$$\pi_i(x_1, x_2, \ldots) := x_i .$$

Since $\pi_i^{-1}(A_i) = S_1 \times \cdots \times S_{i-1} \times A_i \times S_{i+1} \times \cdots$, π_i is an $\mathcal{A}_\otimes$-$\mathcal{A}_i$-measurable mapping. Moreover, $A_1 \times A_2 \times \cdots = \pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) \cap \cdots$, therefore we may characterize the product σ-algebra as the σ-algebra generated by the projection mappings:

$$\mathcal{A}_\otimes = \sigma(\pi_i, i \geq 1) .$$

Example (Euclidean spaces)

The σ-algebra $\mathcal{B}^d$ in $\mathbb{R}^d$, $2 \leq d < \infty$, can be regarded as either a Borel σ-algebra (thus, generated by the open sets) or as a product σ-algebra, because on $\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k}$, $d = d_1 + \cdots + d_k$, it holds that

$$\mathcal{B}^d = \mathcal{B}^{d_1} \otimes \cdots \otimes \mathcal{B}^{d_k} . \tag{2.2}$$

To *prove* this we first note that every open set $O \subset \mathbb{R}^d$ is a countable union of measurable rectangles, e.g., as above,

$$O = \bigcup \left\{[a, b) : [a, b) \subset O\,,\ a, b \in \mathbb{Q}^d\right\} .$$

Thus O belongs to the product σ-algebra. Since $\mathcal{B}^d$ is the smallest σ-algebra including all open sets, it follows that $\mathcal{B}^d \subset \mathcal{B}^{d_1} \otimes \cdots \otimes \mathcal{B}^{d_k}$. Conversely, the projection mappings $\pi_i : \mathbb{R}^d \to \mathbb{R}^{d_i}$ are continuous and thus $\mathcal{B}^d$-$\mathcal{B}^{d_i}$-measurable, and therefore $\mathcal{B}^{d_1} \otimes \cdots \otimes \mathcal{B}^{d_k} = \sigma(\pi_1, \ldots, \pi_k) \subset \mathcal{B}^d$.

Example (The extended real line)

When considering suprema and infima of countably many measurable real functions it is convenient to extend the range to $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$. We equip $\bar{\mathbb{R}}$ with the σ-algebra

$$\bar{\mathcal{B}} := \left\{B \subset \bar{\mathbb{R}} \mid B \cap \mathbb{R} \text{ is a Borel set in } \mathbb{R}\right\} ,$$

called the *Borel σ-algebra* in $\bar{\mathbb{R}}$ (cf. Exercise 2.6), and $\bar{\mathbb{R}}^d$ with the product σ-algebra $\bar{\mathcal{B}}^d$. Here, d is either a natural number, or $d = \infty$. The functions

$$\sup : \bar{\mathbb{R}}^d \to \bar{\mathbb{R}}\,, \qquad \inf : \bar{\mathbb{R}}^d \to \bar{\mathbb{R}}\,,$$

which to every finite or infinite sequence $x_1, x_2, \ldots$ assign its infimum and supremum, respectively, then become $\bar{\mathcal{B}}^d$-$\bar{\mathcal{B}}$-measurable. This follows from

$$\sup^{-1}\big([-\infty, x]\big) = [-\infty, x] \times [-\infty, x] \times \cdots ,$$
$$\inf^{-1}\big([x, \infty]\big) = [x, \infty] \times [x, \infty] \times \cdots ,$$

the measurability criterion, and from the fact that $\bar{\mathcal{B}}$ (similarly to the Borel σ-algebra on the real axis) is generated by the intervals $[-\infty, x]$, and just as well by the intervals $[x, \infty]$.

As a price to be paid, however, it is no longer possible to subtract and divide arbitrary elements of $\bar{\mathbb{R}}$ without entangling oneself in contradictions.

No difficulties arise with the rules

$$\infty + \infty := \infty\,, \quad 0 \cdot \infty := 0\,, \quad a \cdot \infty := \infty \text{ for } a > 0 \quad , (-1) \cdot \infty = -\infty\,;$$

we will use them in the sequel. In contrast, one has to avoid the expressions

$$\infty - \infty\,, \quad \frac{\infty}{\infty}\,;$$

they are (and remain) undefined.

Product σ-algebras have the important property that mappings into a Cartesian product are measurable if and only if the same is true for all their components.

Proposition 2.6. *Let* $(S, \mathcal{A})$ *be a measurable space, let* $\varphi_i : S \to S_i$ *be mappings,* $i \geq 1$. *Then the mapping* $\varphi := (\varphi_1, \varphi_2, \ldots)$ *from* S *to* $S_\times$ *is* $\mathcal{A}$-$\mathcal{A}_\otimes$*-measurable if and only if all mappings* φ_i *are* $\mathcal{A}$-$\mathcal{A}_i$*-measurable.*

Proof. This is a special case of the preceding proposition, as $\varphi_i = \pi_i \circ \varphi$. □

Real Functions

We summarize:

Proposition 2.7. *If* $(S, \mathcal{A})$, $(S_i, \mathcal{A}_i)$, $i \geq 1$, $(S', \mathcal{A}')$ *are measurable spaces, and if the mappings* $\varphi_i : S \to S_i$ *are* $\mathcal{A}$-$\mathcal{A}_i$*-measurable and* $\psi : S_1 \times S_2 \times \cdots \to S'$ *is* $\mathcal{A}_\otimes$-$\mathcal{A}'$*-measurable, then* $\psi \circ (\varphi_1, \varphi_2, \ldots)$ *is* $\mathcal{A}$-$\mathcal{A}'$*-measurable.*

Using this result one may ascertain the measurability of many mappings and sets. We demonstrate this for the particularly important case of functions with values in $\mathbb{R}$ and $\bar{\mathbb{R}} = [-\infty, \infty]$ ($\mathbb{R}^d$ and $\bar{\mathbb{R}}$ are equipped with the Borel σ-algebra $\mathcal{B}^d$ resp. $\bar{\mathcal{B}}$).

Here, the simplest functions are the characteristic functions 1_A of subsets A of S, taking the value 1 on A and 0 on A^c. 1_A is a measurable function if and only if A is a measurable subset.

Let now $f_1, f_2 : S \to \mathbb{R}$ be measurable functions and let $c_1, c_2 \in \mathbb{R}$. Then the linear combination $c_1 f_1 + c_2 f_2$ is a measurable function. This follows from the representation

$$c_1 f_1 + c_2 f_2 = \varphi \circ (f_1, f_2) ,$$

where $\varphi(x, y) := c_1 x + c_2 y$, due to continuity, is a Borel measurable mapping from $\mathbb{R}^2$ to $\mathbb{R}$. In the same way one obtains the measurability of

$$f_1 \cdot f_2 , \quad \max(f_1, f_2) , \quad \min(f_1, f_2)$$

and, for every measurable f, the measurability of

$$f^+ := \max(f, 0) , \quad f^- := \max(-f, 0) , \quad |f| = f^+ + f^- .$$

For measurable functions $f_1, f_2 : S \to \mathbb{R}$ the measurability of the set

$$\{f_1 = f_2\} := \big\{x \in S : f_1(x) = f_2(x)\big\}$$

results from the fact that $\{f_1 = f_2\} = (f_1, f_2)^{-1}(D)$ with the "diagonal" $D := \{(x, y) \in \mathbb{R}^2 : x = y\}$, since D, being a closed subset of $\mathbb{R}^2$, is Borel measurable.

Analogously one obtains the measurability of sets like

$$\{f_1 \le f_2\} := \big\{x \in S : f_1(x) \le f_2(x)\big\}$$

or $\{f_1 \neq f_2\}$, $\{f_1 < f_2\}$.

In the same manner one may construct new measurable functions from infinite sequences $f_1, f_2, \dots$ of given measurable functions from S to $\mathbb{R}$, extending $\mathbb{R}$ in the process to $\bar{\mathbb{R}}$ when necessary. We have shown that the mappings $\sup, \inf : \bar{\mathbb{R}}^\infty \to \bar{\mathbb{R}}$ are measurable; therefore, for measurable $f_1, f_2, \dots$ their pointwise supremum and infimum

$$\sup_{n\ge1} f_n = \sup\circ(f_1, f_2, \dots) , \quad \inf_{n\ge1} f_n = \inf\circ(f_1, f_2, \dots)$$

are measurable. As a consequence, the pointwise limit superior and limit inferior

$$\limsup_{n\to\infty} f_n = \inf_{m\ge1} \sup_{n\ge m} f_n , \quad \liminf_{n\to\infty} f_n = \sup_{m\ge1} \inf_{n\ge m} f_n,$$

are measurable. Moreover, $\{\lim_n f_n \text{ exists}\}$ is a measurable set, since

$$\{\lim_n f_n \text{ exists}\} = \{\limsup_n f_n = \liminf_n f_n\} \cap \{-\infty < \limsup_n f_n < \infty\} .$$

If the sequence $f_1, f_2, \dots$ converges pointwise, we have $\lim_n f_n = \limsup_n f_n$, and thus $\lim_n f_n$ is a measurable function. With that property of measurable mappings we are already acquainted.

For the theory of integration, the following characterization of measurable nonnegative functions will be important. Using it we will extend properties of the integral from specific sets of measurable functions to all measurable functions. *Nonnegative functions* are always understood as functions values in $\bar{\mathbb{R}}_+ = [0, \infty]$.

Proposition 2.8 (Monotonicity principle). *Let* $(S, \mathcal{A})$ *be a measurable space and let* $\mathcal{K}$ *be a set of functions* $f : S \to \bar{\mathbb{R}}_+$ *with the properties*

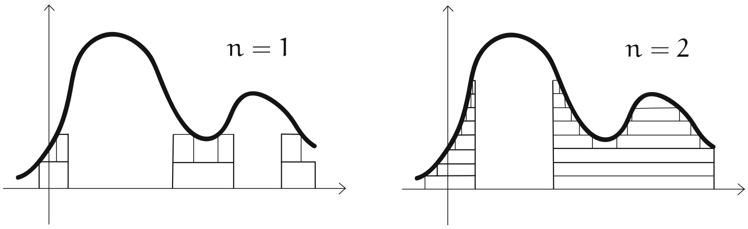
(i) $f_1, f_2 \in \mathcal{K}\,,\ c_1, c_2 \in \mathbb{R}_+ \quad \Rightarrow \quad c_1f_1 + c_2f_2 \in \mathcal{K}$,
(ii) $f_1, f_2, \dots \in \mathcal{K}\,, f_1 \le f_2 \le \cdots \quad \Rightarrow \quad \sup_n f_n \in \mathcal{K}$,
(iii) $1_A \in \mathcal{K}$ *for all* $A \in \mathcal{A}$.

Then $\mathcal{K}$ *includes all nonnegative measurable functions on* S (*with values in* $\bar{\mathbb{R}}_+$*, according to our terminology*).

Proof. Let $f : S \to \bar{\mathbb{R}}_+$ be measurable. The sets $A_{k,n} := \{k2^{-n} < f \le (k+1)2^{-n}\}$ belong to $\mathcal{A}$, for all natural numbers k and n. The functions

$$f_n := \sum_{k=1}^{n2^n} \frac{k}{2^n} 1_{A_{k,n}} + n\, 1_{\{f=\infty\}}$$

then belong to $\mathcal{K}$ by virtue of (i) and (iii).



One has that $f_1 \le f_2 \le \cdots$ and $\sup_{n\ge1} f_n = f$, therefore using (ii) we obtain $f \in \mathcal{K}$, as asserted. □

Exercises

2.1 Let S be a set. Which σ-algebra is generated by the subsets of S consisting of a single element only ? What are the measurable functions $f : S \to \mathbb{R}$?

2.2 Let $E_1, E_2, \ldots$ be a partition of S, that is, a sequence of disjoint subsets of S with $\bigcup_{n\ge1} E_n = S$. Let $\mathcal{A}$ be the σ-algebra generated by those sets. Describe all sets which belong to $\mathcal{A}$.

2.3 Let $\mathcal{A}_1, \mathcal{A}_2$ be σ-algebras on S. Is $\mathcal{A}_1 \cap \mathcal{A}_2$ a σ-algebra? What about $\mathcal{A}_1 \cup \mathcal{A}_2$?
Hint: One may construct counterexamples from σ-algebras with 4 elements.

2.4 Prove: The σ-algebra $\bar{\mathcal{B}}$ on $\bar{\mathbb{R}}$ is generated by the intervals $[-\infty, b]$, $b \in \mathbb{R}$.

2.5 Let S be a metric space with metric d. Prove:
(i) Every closed set $F \subset S$ can be obtained as an intersection of countably many open sets (one says that F is a G_δ-set).
(ii) The Borel σ-algebra in S equals the smallest system $\mathcal{B}'$ of sets which includes all open sets and moreover, for every sequence $B_1, B_2, \ldots$, the sets $\bigcup_{n\ge1} B_n$ and $\bigcap_{n\ge1} B_n$.
Hint: Consider the system $\{B \in \mathcal{B}' : B^c \in \mathcal{B}'\}$.

2.6 Let $m : \bar{\mathbb{R}} \to \mathbb{R}$ be strictly monotone and bounded. Prove that $\bar{d}(x, y) := |m(x) - m(y)|$ defines a metric $\bar{d}$ on $\bar{\mathbb{R}}$ and that the corresponding Borel σ-algebra equals $\bar{\mathcal{B}}$.
Hint: The system $O \subset \bar{\mathbb{R}}$ of open sets depends upon whether and where m has jumps!

2.7 **The graph of a measurable mapping** Let $\varphi, \psi, \psi' : S \to S'$ be $\mathcal{A}$-$\mathcal{A}'$-measurable mappings and assume that the "diagonal" $D := \{(x, y) \in S' \times S' : x = y\}$ belongs to $\mathcal{A}' \otimes \mathcal{A}'$. Prove that $\{\psi = \psi'\} \in \mathcal{A}$ and conclude that

$$\big\{(x, y) \in S \times S' : y = \varphi(x)\big\} \in \mathcal{A} \otimes \mathcal{A}' .$$

2.8 Let S be uncountable and set $\mathcal{A} := \{A \subset S : A \text{ or } A^c \text{ is countable}\}$. Prove:
(i) $\mathcal{A}$ is a σ-algebra.

(ii) For every $A' \in \mathcal{A} \otimes \mathcal{A}$, either A' or $(A')^c$ is thin. Here we say that $A' \subset S^2$ is “thin”, if $A' \subset (A \times S) \cup (S \times A)$ for some countable set $A \subset S$.

(iii) The diagonal $D := \{(x, y) \in S \times S \,:\, x = y\}$ does not belong to $\mathcal{A} \otimes \mathcal{A}$.

2.9 A function $g : \mathbb{R}^d \to \bar{\mathbb{R}}$ is called upper semicontinuous, if

$$\limsup_{y \to x} g(y) \;\leq\; g(x)$$

holds for all $x \in \mathbb{R}^d$. Prove:

(i) g is upper semicontinuous if and only if the sets $\{g < a\} := \{x \in \mathbb{R}^d : g(x) < a\}$ are open for all real numbers a.

(ii) Upper semicontinuous functions are Borel measurable.

(iii) For every (not necessarily measurable) function $f : \mathbb{R}^d \to \mathbb{R}$, the functions

$$g(x) \;:=\; \lim_{\varepsilon \downarrow 0} \sup_{|y-x| \leq \varepsilon} f(y)\;,\;\; h(x) \;:=\; \lim_{\varepsilon \downarrow 0} \inf_{|y-x| \leq \varepsilon} f(y)\;, \quad x \in \mathbb{R}\;,$$

are upper semicontinuous, resp. lower semicontinuous (that is, $-h$ is upper semicontinuous). Prove: The set $C \subset \mathbb{R}^d$ of points of continuity of f is a Borel set, and $f1_C$ is Borel measurable.

(iv) A function $f : \mathbb{R}^d \to \mathbb{R}$ having at most countably many points of discontinuity is Borel measurable.

Measures

3

Carrying measures is an essential purpose of measurable spaces.

Definition

Let $(S, \mathcal{A})$ be a measurable space. A mapping μ which to every $A \in \mathcal{A}$ assigns a number $\mu(A) \geq 0$, or possibly the value $\mu(A) = \infty$, is called a *measure*, if:

(i) $\mu(\emptyset) = 0$,

(ii) σ-additivity: $\mu\bigl(\bigcup_{n\geq 1} A_n\bigr) = \sum_{n\geq 1} \mu(A_n)$ for every finite or infinite sequence $A_1, A_2, \ldots$ of pairwise disjoint measurable sets.

The triple $(S, \mathcal{A}, \mu)$ is called a *measure space*. If $\mu(S) = 1$, then μ is called a *probability measure*. More generally, μ is called *finite* if $\mu(S) < \infty$, and σ*-finite* if there exist measurable sets $A_1 \subset A_2 \subset \cdots$ such that $\bigcup_{n\geq 1} A_n = S$ and $\mu(A_n) < \infty$ holds for all n.

In the Introduction we have been guided by the idea that $\mu(A)$ is the volume of A. One also may think of μ as describing a mass distribution on S, and then $\mu(A)$ equals the mass of A. In probability theory one interprets the elements A of the σ-algebra as observable events occuring with probability $\mu(A)$.

σ-finite measures are of interest for two reasons. Firstly, some important measures are σ-finite, e.g., the Lebesgue measure, on $\mathbb{R}^d$ which we will address soon. Secondly, properties of finite measures often extend to the σ-finite case. One achieves this by replacing a given σ-finite measure μ with finite measures $\mu_n(\cdot) := \mu(\cdot \cap A_n)$ and then passing to the limit $n \to \infty$. Often this does not present any difficulties whatsoever, so that one may omit the details.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_3

Example

1. A *Dirac measure*[1] is a probability measure whose total mass is concentrated in a single point. The Dirac measure δ_x in the point $x \in S$ of a measurable space is defined as

$$\delta_x(A) \ := \begin{cases} 1\,, & \text{if } x \in A\,, \\ 0\,, & \text{if } x \notin A\,. \end{cases}$$

   As values it takes on only 0 and 1.
2. A measure μ is called *discrete* if its total mass is concentrated in a countable measurable set, that is, if $\mu(C^c) = 0$ holds for some countable set $C \subset S$. In this case, μ is specified by its *weights* $\mu_x := \mu\big(\{x\}\big)$, $x \in C$, according to the formula

$$\mu(A) \ = \sum_{x \in A \cap C} \mu_x\,.$$

   Conversely, from any family $(\mu_x)_{x\in C}$ of nonnegative numbers, using this formula one obtains a discrete measure μ.

The following proposition summarizes several essential properties of measures. For sets $A, A_1, A_2, \ldots \subset S$ we write

$$A_n \uparrow A\,, \quad \text{if } A_1 \subset A_2 \subset \cdots \text{ and } A = \bigcup_{n\geq 1} A_n\,,$$

$$A_n \downarrow A\,, \quad \text{if } A_1 \supset A_2 \supset \cdots \text{ and } A = \bigcap_{n\geq 1} A_n\,.$$

Proposition 3.1. *For any measure* μ *und any measurable sets* $A, A_1, A_2, \ldots$ *there holds:*

(i) Monotonicity: $\mu(A_1) \leq \mu(A_2)$, *if* $A_1 \subset A_2$,
(ii) σ-subadditivity: $\mu\big(\bigcup_{n\geq 1} A_n\big) \leq \sum_{n\geq 1} \mu(A_n)$,
(iii) σ-continuity: *If* $A_n \uparrow A$, *then* $\mu(A_n) \to \mu(A)$ *for* $n \to \infty$.
If $A_n \downarrow A$ *and moreover* $\mu(A_1) < \infty$, *then* $\mu(A_n) \to \mu(A)$ *for* $n \to \infty$ *as well.*

Proof. (i) In the case $A_1 \subset A_2$, A_2 equals the disjoint union of A_1 and $A_2 \setminus A_1$, and $\mu(A_1) \leq \mu(A_1) + \mu(A_2 \setminus A_1) = \mu(A_2)$ follows by additivity.

[1] PAUL DIRAC, 1902–1984, born in Bristol, active in Cambridge. He is famous in particular for his contributions to the foundations of quantum mechanics. In 1933 he was awarded the Nobel prize for physics.

(ii) To begin with, we have $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2 \setminus A_1) \le \mu(A_1) + \mu(A_2)$ due to additivity und monotonicity. For finite unions it follows by induction that $\mu(A_1 \cup \dots \cup A_k) \le \mu(A_1 \cup \dots \cup A_{k-1}) + \mu(A_k) \le \mu(A_1) + \dots + \mu(A_{k-1}) + \mu(A_k)$. It remains to prove the assertion for infinite unions. For this, one passes in $\mu(A_1 \cup \dots \cup A_k) \le \sum_{n\ge 1} \mu(A_n)$ to the limit $k \to \infty$, using the σ-continuity of measures, which we will prove next.

(iii) Assuming $A_n \uparrow A$, the sets $A'_1 := A_1$, $A'_k := A_k \setminus A_{k-1}$, $k \ge 2$, are disjoint and we have $A_n = \bigcup_{k=1}^n A'_k$, $A = \bigcup_{k=1}^\infty A'_k$. Consequently, as $n \to \infty$,

$$\mu(A_n) = \mu\Big(\bigcup_{k=1}^n A'_k\Big) = \sum_{k=1}^n \mu(A'_k) \;\to\; \sum_{k=1}^\infty \mu(A'_k) = \mu\Big(\bigcup_{k=1}^\infty A'_k\Big) = \mu(A)\,.$$

This yields the first assertion. Assuming $A_n \downarrow A$ we get $A''_n \uparrow A_1 \setminus A$ for $A''_n := A_1 \setminus A_n$, $n \ge 1$. Consequently,

$$\mu(A_n) + \mu(A''_n) = \mu(A_1) = \mu(A) + \mu(A_1 \setminus A)\,.$$

Passing to the limit $n \to \infty$ yields the second assertion, using the first assertion as well as the assumption $\mu(A_1) < \infty$.

□

► **Remark** The condition $\mu(A_1) < \infty$ in the last assertion cannot be omitted without replacement. A counterexample is provided by the sequence of sets $A_n := \{m \in \mathbb{N} : m \ge n\}$. The A_n all have measure ∞ for the counting measure μ on $\mathbb{N}$ defined by $\mu(A) := \#A$. On the other hand, $\bigcap_{n\ge 1} A_n = \emptyset$ has measure zero.

Measures can be mapped to other measure spaces via measurable mappings. This issue will become important presently.

Definition

Let $(S, \mathcal{A})$, $(S', \mathcal{A}')$ be measurable spaces, let $\varphi : S \to S'$ be measurable, and let μ be a measure on $\mathcal{A}$. The measure μ' on S given by

$$\mu'(A') := \mu\big(\varphi^{-1}(A')\big)\,, \quad A' \in \mathcal{A}'\,,$$

is called the *image measure* of μ *under the mapping* φ. We write $\mu' = \varphi(\mu)$.

A short computation shows that μ' indeed is a measure: $\mu'(\emptyset) = \mu(\emptyset) = 0$ and $\mu'\big(\bigcup_{n\ge 1} A'_n\big) = \mu\big(\bigcup_{n\ge 1} \varphi^{-1}(A'_n)\big) = \sum_{n\ge 1} \mu\big(\varphi^{-1}(A'_n)\big) = \sum_{n\ge 1} \mu'(A'_n)$ for pairwise disjoint sets $A'_1, A'_2, \ldots \in \mathcal{A}'$. Just as quickly one convinces oneself that

$$(\psi \circ \varphi)(\mu) = \psi(\varphi(\mu))\,.$$

From $\varphi(x) = y$ it follows that $\varphi(\delta_x) = \delta_y$, we thus have transferred φ to measures in a canonical manner.

Null Sets

We now broach the topic of those measurable sets which a given measure does not distinguish from the empty set.

Definition (Null set)

Let $(S, \mathcal{A}, \mu)$ be a measure space. A set $A \subset S$ is called a *null set*, more precisely a μ-*null set*, if $A \in \mathcal{A}$ and $\mu(A) = 0$.

The system $\mathcal{N} \subset \mathcal{A}$ of all null sets of a not identically vanishing measure μ has the following properties, as a consequence of monotonicity and σ-subadditivity:

$$\emptyset \in \mathcal{N}\,, \quad S \notin \mathcal{N}\,,$$
$$A \in \mathcal{N}, A' \in \mathcal{A}, A' \subset A \quad \Rightarrow \quad A' \in \mathcal{N}\,,$$
$$A_1, A_2, \ldots \in \mathcal{N} \quad \Rightarrow \quad \bigcup_{n \geq 1} A_n \in \mathcal{N}\,.$$

If a property holds for all elements of S except for elements in some null set, then one says that the property holds *almost everywhere*.

Definition

Let $(S, \mathcal{A}, \mu)$ be a measure space. Two measurable mappings $\varphi, \psi : S \to S'$ are called *equal almost everywhere*, more precisely *equal* μ-*almost everywhere*, if $\{\varphi \neq \psi\}$ is a null set. We write

$$\varphi = \psi \ \text{ a.e.}$$

and say that $\varphi(x) = \psi(x)$ for μ-almost all x.

In probability theory one speaks of *almost sure equality*. It is an equivalence relation. In the same manner one writes

$$\varphi \leq \psi \ \text{ a.e.} \quad :\Leftrightarrow \quad \{\varphi > \psi\} \text{ is a null set}\,,$$

whenever the range S' of φ and ψ is equipped with an order relation $\leq$.
Null sets will become important for us specifically in the context of convergence.

Definition

Let $(S, \mathcal{A}, \mu)$ be a measure space, let (S', d') be a metric space, and let $\varphi, \varphi_1, \varphi_2, \ldots$ be measurable mappings. We say that φ_n *converges to* φ *almost everywhere*, and write

$$\varphi_n \to \varphi \text{ a.e.} ,$$

if $\{\varphi_n \not\to \varphi\} := \{x \in S : \varphi_n(x) \not\to \varphi(x)\}$ is a null set.

► **Remark** For any measure space $(S, \mathcal{A}, \mu)$, the system

$$\tilde{\mathcal{A}} := \left\{\tilde{A} \subset S \;:\; \exists\, A_1, A_2 \in \mathcal{A} \text{ with } A_1 \subset \tilde{A} \subset A_2 \text{ and } \mu(A_2 \setminus A_1) = 0\right\}$$

is a σ-algebra in S which contains $\mathcal{A}$. One may also describe it as the σ-algebra generated by $\mathcal{A} \cup \tilde{\mathcal{N}}$, where $\tilde{\mathcal{N}}$ denotes the system of all subsets of all null sets. Moreover, for any $\tilde{A} \in \tilde{\mathcal{A}}$,

$$\tilde{\mu}(\tilde{A}) := \mu(A_1) = \mu(A_2) ,$$

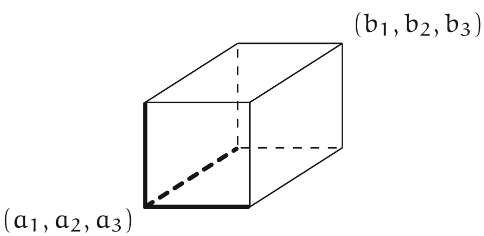
is well defined. $\tilde{\mu}$ is a measure which extends μ to $\tilde{\mathcal{A}}$. The measure space $(S, \tilde{\mathcal{A}}, \tilde{\mu})$ is called the *completion* of $(S, \mathcal{A}, \mu)$. (Proof as exercise)

The Lebesgue Measure on $\mathbb{R}^d$

Now we will see that the concept of a measure on a σ-algebra works out well in an especially important case. The following nontrivial result says that there exists a unique measure on the Borel σ-algebra $\mathcal{B}^d$ on $\mathbb{R}^d$ (d finite) which associates to every d-dimensional interval its "natural" volume. One usually considers half-open intervals

$$[a, b) := [a_1, b_1) \times \cdots \times [a_d, b_d) ,$$

where $a = (a_1, \ldots, a_d)$, $b = (b_1, \ldots, b_d) \in \mathbb{R}^d$. Half-open intervals have the advantage that they cover the whole space $\mathbb{R}^d$ completely and without intersections. In the following picture, the bold edges belong to the corresponding interval.



Proposition 3.2. *On the Borel sets of* $\mathbb{R}^d$ *there exists a unique measure, denoted by* λ^d*, with the property that for all* $a_1 < b_1, \ldots, a_d < b_d$

$$\lambda^d\big([a,b)\big) = (b_1 - a_1)\cdots(b_d - a_d)\ ,$$

where $a = (a_1, \ldots, a_d)$ *and* $b = (b_1, \ldots, b_d)$.

We postpone the proof, we will show uniqueness in Chap. 7 and existence in Chap. 11. λ^d is called the *Lebesgue measure on* $\mathcal{B}^d$ (one also speaks of the *Lebesgue-Borel measure*). Its completion, too, is termed Lebesgue measure. In the case $d = 1$, we more briefly write $\mathcal{B}$ and λ for the Borel σ-algebra $\mathcal{B}^1$ and the measure λ^1.

We treat some important properties of the Lebesgue measure.

Proposition 3.3. *The Lebesgue measure* λ^d *on* $\mathcal{B}^d$ *is the only measure on* $\mathcal{B}^d$ *satisfying the following two properties:*

(i) Translation invariance: $\lambda^d(B) = \lambda^d(B')$, *if* $B, B' \in \mathcal{B}^d$ *map into each other by translation.*
(ii) Normalization: $\lambda^d\big([0,1)^d\big) = 1$ for the d-dimensional unit cube $[0,1)^d$.

Proof. Only (i) has to be proved. We consider, for fixed $v \in \mathbb{R}^d$, the translation mapping $x \mapsto \varphi(x) := x + v$ on $\mathbb{R}^d$ and the image measure $\mu := \varphi(\lambda^d)$. Translation maps intervals to intervals of equal measure, that is, it holds that $\mu\big([a,b)\big) = (b_1 - a_1)\cdots(b_d - a_d)$. Thus μ satisfies the property which characterizes the Lebesgue measure. It follows that $\mu = \lambda^d$, therefore $\lambda^d(B) = \lambda^d\big(\varphi^{-1}(B)\big)$, yielding assertion (i).

Conversely, let μ be any measure satisfying (i) and (ii). Then for every natural number n we have

$$\mu\big([0,n^{-1})^d\big) = n^{-d}\ ,$$

since the cube $[0,1)^d$ decomposes in n^d subcubes, which all arise from $[0,n^{-1})^d$ by translation and therefore have identical measure, by the assumption. From such cubes one can compose all those half-open d-dimensional intervals whose boundaries a and b consist of rational components only. By additivity,

$$\mu\big([a,b)\big) = (b_1 - a_1)\cdots(b_d - a_d)$$

follows for rational numbers $a_i < b_i$. Because the rational numbers form a dense subset of the real numbers, we may enclose, from outside as well as from inside, arbitrary intervals by intervals with rational vertices. The latter formula then follows for arbitrary $a_i < b_i$, by the monotonicity property of measures. Therefore, μ enjoys the property characterizing the Lebesgue measure, and so $\mu = \lambda^d$. □

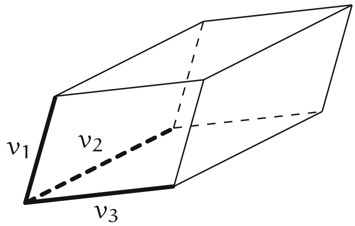
Proposition 3.4. *For the Lebesgue measure it holds that:*

(i) $\lambda^d(H) = 0$ *for every hyperplane* $H \subset \mathbb{R}^d$.
(ii) *If* $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ *is linear, then* $\varphi(B)$ *is a Borel set for every Borel set* $B \subset \mathbb{R}^d$, *and*

$$\lambda^d\big(\varphi(B)\big) = |\det \varphi| \cdot \lambda^d(B) \ .$$

Phrased in a different manner: If φ maps the canonical unit vectors $e_1, \ldots, e_d$ to the vectors $v_1, \ldots, v_d \in \mathbb{R}^d$, then it maps the unit cube $[0,1)^d$ to the (half-open) parallelepiped spanned by the vectors $v_1, \ldots, v_d$:

$$P[v_1, \ldots, v_d] := \Big\{ \sum_{i=1}^{d} c_i v_i \in \mathbb{R}^d \ : \ 0 \le c_i < 1 \ , \ i = 1, \ldots, d \Big\} \ .$$



Since $\det \varphi = \det[v_1, \ldots, v_d]$, we get

$$\lambda^d\big(P[v_1, \ldots, v_d]\big) = \big|\det[v_1, \ldots, v_d]\big| \ .$$

This is not surprising since the determinant $\det[v_1, \ldots, v_d]$ can be interpreted as the *oriented* volume of a parallelepiped, as we know from linear algebra. Except for the orientation (the sign of the determinant), measure theory thus yields the same result.

Proof. (i) Every hyperplane H can be covered by countably many sets which arise by translation from a single $(d-1)$-dimensional rectangle Q spanned by some orthogonal vectors $b_2, \ldots, b_d$. Let b_1 be orthogonal to $b_2, \ldots, b_d$; the sets $Q + rb_1$, $0 \le r \le 1$, are pairwise disjoint and have identical Lebesgue measure due to translation invariance. That measure has to be equal to zero, because otherwise the rectangle spanned by $b_1, \ldots, b_d$ would have infinite measure. Thus Q, and therefore H too, has measure 0.

(ii) The case $\det\varphi = 0$ is already covered by (i). So let us assume that φ has an inverse ψ which, being a linear mapping, is continuous and therefore Borel measurable. We conclude that $\varphi(B) = \psi^{-1}(B)$ is Borel measurable whenever B is a Borel set.

The bijectivity of φ has additional consequences: $\mu(\cdot) := \lambda^d(\varphi(\cdot))$ defines a measure. Since $\varphi(B + v) = \varphi(B) + \varphi(v)$ we have $\mu(B + v) = \mu(B)$ for every $v \in \mathbb{R}^d$; thus μ is translation invariant. Moreover, $0 < c < \infty$ holds for $c := \mu\big([0,1)^d\big)$, as $\varphi\big([0,1)^d\big)$ contains small cubes and is contained in a large cube. The characterization of the Lebesgue measure in the previous proposition now yields $\mu = c\lambda^d$.

It remains to determine the value of c. We begin by considering two simple cases:

First, let σ be a linear mapping having the unit vectors $e_1, \dots, e_d$ as eigenvectors, with eigenvalues $\varepsilon_1, \dots, \varepsilon_d > 0$. Then $[0,1)^d$ transforms into the interval $[0,\varepsilon_1) \times \cdots \times [0,\varepsilon_d)$, and we directly obtain $c = \varepsilon_1 \cdots \varepsilon_d$, which in turn equals $\det\sigma$.

Secondly, let τ be an orthogonal mapping. Then τ maps the unit ball B onto itself. Since B can be sandwiched between cubes, $0 < \lambda^d(B) < \infty$. In this case we therefore have $c = 1$; on the other hand, the determinant of an orthogonal mapping is known to be ± 1.

The assertion now results from the fact that every linear mapping φ can be represented as $\varphi = \tau_1 \circ \sigma \circ \tau_2$, where σ is as above and τ_1, τ_2 are orthogonal mappings ("singular value decomposition"). Indeed, the assertion follows from the special cases considered above and known properties of determinants:

$$\lambda^d\big(\varphi(B)\big) = \lambda^d\big(\tau_1(\sigma(\tau_2(B)))\big) = \big|\det\tau_1 \det\sigma \det\tau_2\big|\lambda^d(B) = |\det\varphi|\lambda^d(B)\ .$$

(In Exercise 3.9, we will recap the singular value decomposition of matrices.)

□

Exercises

3.1 Let $A = \bigcup_{n\geq 1} A_n$ and $A' = \bigcup_{n\geq 1} A'_n$. Check whether for every measure μ it is true that

$$\mu(A \setminus A') \leq \sum_{n\geq 1} \mu(A_n \setminus A'_n)\ , \quad \mu(A\Delta A') \leq \sum_{n\geq 1} \mu(A_n \Delta A'_n)\ .$$

3.2 Let $\mu_1 \leq \mu_2 \leq \cdots$ be a sequence of measures on a σ-algebra, that is, $\mu_1(A) \leq \mu_2(A) \leq \cdots$ holds for every measurable set A. Prove that $\mu(A) := \lim_n \mu_n(A)$ defines a measure μ.

3.3 Let $\mathcal{A}$ be the system of all sets $A \subset \mathbb{N}$ for which the limit

$$\iota(A) := \lim_{n\to\infty} \frac{1}{n}\#(A \cap \{1,2,\dots,n\})$$

exists. Prove: (i) ι is additive, but not σ-additive, (ii) $\mathcal{A}$ is not a σ-algebra.

3.4 **Existence of non-measurable sets, due to Vitali** Let $N \subset [0,1]$ a set having the property that for every real number a there exists a unique number $b \in N$ such that $a - b$ is rational. Prove:

(i) $N + r$ and $N + r'$ are disjoint for any rational numbers $r \neq r'$.

(ii) $[0,1] \subset \bigcup_{r \in \mathbb{Q} \cap [-1,1]} (N + r) \subset [-1, 2]$.

(iii) N is not a Borel set.

Remark: N is a complete set of representatives for the equivalence relation on $\mathbb{R}$ given by $a \sim b :\Leftrightarrow a - b \in \mathbb{Q}$. One obtains N using the axiom of choice from set theory.

3.5 **Egorov's theorem** Let μ be a finite measure, and let $f_1, f_2, \dots$ converge μ-a.e. to f. We want to prove that for each $\varepsilon > 0$ there exists a measurable set $A \subset S$ such that $f_1, f_2, \dots$ converges uniformly to f on A and that $\mu(A^c) \leq \varepsilon$. For this purpose, prove:

(i) Let $\delta > 0$ and $A'_m := \bigcup_{n \geq m} \{|f_n - f| > \delta\}$. Then $\bigcap_{m \geq 1} A'_m \subset \{f_n \not\to f\}$ and $\mu\big(A'_m\big) \to 0$ for $m \to \infty$.

(ii) For each $\varepsilon > 0$ there exist natural numbers $m_1 < m_2 < \cdots$ such that $\mu(A_k) \leq \varepsilon 2^{-k}$ for $A_k := \bigcup_{n \geq m_k} \{|f_n - f| > 1/k\}$.

(iii) f converges uniformly on $A := \bigcap_{k \geq 1} A_k^c$, and $\mu(A^c) \leq \varepsilon$.

3.6 On $\mathbb{R}$, we consider the Borel measurable functions $f = 1_{\mathbb{Q}}$ and $g = 1_{[0,1]}$. Which of these functions are (i) a.e. continuous, (ii) a.e. equal to a continuous function (with respect to the Lebesgue measure)?

3.7 Let $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ be linear and bijective. Prove that for the image of the Lebesgue measure under φ it holds that $\varphi(\lambda^d)(\cdot) = |\det \varphi|^{-1} \lambda^d(\cdot)$.

3.8 Let $B \subset [0,1)$ be a Borel set. Prove that for each $\varepsilon > 0$ there exist (half-open) intervals $I_1, \dots, I_k \subset [0,1)$ such that $\lambda^1\big(B \Delta \bigcup_{j=1}^k I_j\big) < \varepsilon$. In addition, consider the d-dimensional case.

Hint: Consider the system of all sets B having the stated property.

3.9 **Singular value decomposition** Let M be an invertible $d \times d$-matrix, and let M^* be its adjoint. Prove:

(i) M^*M is selfadjoint and invertible, with strictly positive eigenvalues $\varepsilon_1^2, \dots, \varepsilon_d^2$. Thus there exists an orthogonal matrix O such that $M^*M = O^*D^2O$; here D denotes the diagonal matrix with entries $\varepsilon_1, \dots, \varepsilon_d$.

(ii) The mapping $DOx \mapsto Mx$, $x \in \mathbb{R}^d$, is well-defined, linear and orthogonal, that is, $|DOx|^2 = |Mx|^2$ for all x.

(iii) There exists an orthogonal matrix V such that $M = VDO$ ("singular value decomposition").

4 The Integral of Nonnegative Functions

Given a measure μ on a measurable space $(S, \mathcal{A})$, we now define the integral for arbitrary measurable functions

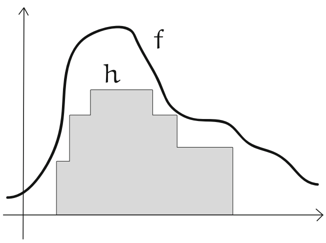
$$f \geq 0 \,.$$

Here we consider measurable functions on S taking values in $\bar{\mathbb{R}}_+ = [0, \infty]$.

The integral will be defined with the aid of *elementary functions*. These are measurable functions $h \geq 0$ having at most finitely many different real values. Thus

$$h = \sum_z z \cdot 1_{\{h=z\}} \,,$$

where the summation runs over the finitely many real function values z of h. In the case $S = \mathbb{R}$ the elementary functions include the step functions, for which the sets $\{h = z\}$ are intervals or finite union of intervals, but additionally, due to



the diversity of the Borel sets, quite different functions which no longer can be represented graphically.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_4

The integral of $f \geq 0$ arises through exhaustion from below by means of elementary functions, as follows:

Definition

Let $f : S \to \bar{\mathbb{R}}_+$ be measurable. The *integral of* f *w.r.t. the measure* μ (more precisely, the *Lebesgue integral*) is defined as

$$\int f \, d\mu \ := \ \sup \Big\{ \sum_z z \cdot \mu(h = z) \ : \ h \geq 0 \text{ is elementary}, \ h \leq f \Big\} .$$

Here, for $\mu\big(\{h = z\}\big)$ we simply write $\mu(h = z)$. The integral may possibly have the value ∞. Sometimes, in particular when the integrand f also depends on other variables in addition to x, one has to specify clearly with respect to which variable one integrates. In that case one writes the integral as

$$\int f(x) \, \mu(dx) .$$

One can interpret the integral as the "content" of the region between 0 and f w.r.t. μ (we will come back to this in Exercise 8.4). In the case of a probability measure the integral may be interpreted as the "mean value" of f w.r.t. μ. If, in particular, μ is a probability measure on $\mathbb{R}^+$ interpreted as a mass distribution, $\int x \, \mu(dx)$ becomes its center of mass. In probability theory one uses integrals in a similar manner, in order to define expectation values.

From the definition we at once draw a simple but important conclusion.

Proposition 4.1 (Markov's Inequality[1]). *Let* $f \geq 0$ *be measurable, and let* z *be a nonnegative number. Then*

$$z \cdot \mu(f \geq z) \leq \int f \, d\mu .$$

Proof. For the elementary function $h := z \cdot 1_{\{f \geq z\}}$ we have $0 \leq h \leq f$. □

The following properties of the integral are immediate consequences of its definition.

[1] Andrei Markov, 1856–1922, born in Ryazan, active in St. Petersburg. He is mainly known for his fundamental contributions to probability theory.

Proposition 4.2. *For arbitrary measurable functions* $f, g \geq 0$ *it holds that:*

(i) $f \leq g$ a.e. $\Rightarrow$ $\int f\,d\mu \leq \int g\,d\mu$,
(ii) $f = g$ a.e. $\Rightarrow$ $\int f\,d\mu = \int g\,d\mu$,
(iii) $\int f\,d\mu = 0$ $\Leftrightarrow$ $f = 0$ a.e.,
(iv) $\int f\,d\mu < \infty$ $\Rightarrow$ $f < \infty$ a.e.

Proof. (i) If $h \geq 0$ is elementary with $h \leq f$, then also $h' := h \cdot 1_{\{f \leq g\}}$ is elementary and $h' \leq g$. By assumption, $\sum_z z \cdot \mu(h' = z) = \sum_z z \cdot \mu(h = z, f \leq g) = \sum_z z \cdot \mu(h = z)$, and the assertion follows from the definition of the integral.

(ii) Follows from (i).

(iii) The implication $\Leftarrow$ follows from (ii) and the definition of the integral. Conversely, let $\int f\,d\mu = 0$. For any $n \in \mathbb{N}$, Markov's inequality yields $\mu(f \geq 1/n) = 0$. Since $\{f \geq 1/n\} \uparrow \{f > 0\}$ and due to σ-continuity, $\mu(f > 0) = 0$ follows, and therefore $f = 0$ a.e.

(iv) From $h := z \cdot 1_{\{f=\infty\}} \leq f$ for all $z > 0$ it follows that $z \cdot \mu(f = \infty) \leq \int f\,d\mu$ for all $z > 0$. From $\int f\,d\mu < \infty$ we therefore get $\mu(f = \infty) = 0$, which yields the assertion. □

The following proposition, also called *Beppo Levi's*[2] *theorem*, is a key result element in the theory of the Lebesgue integral.

Proposition 4.3 (Monotone Convergence Theorem). *Let* $0 \leq f_1 \leq f_2 \leq \cdots$ *hold for measurable functions* $f_1, f_2, \ldots$, *and set* $f := \sup_{n\geq 1} f_n$. *Then*

$$\int f\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu\ .$$

Proof. By Proposition 4.2 (i), the sequence $\int f_n\,d\mu$ increases monotonically, and $\lim_n \int f_n\,d\mu \leq \int f\,d\mu$. To prove the reverse inequality, let $h \geq 0$ be elementary with $h \leq f$ and let $\varepsilon > 0$. The elementary functions

$$h_n := (h - \varepsilon)^+ \cdot 1_{\{f_n > h - \varepsilon\}}$$

[2]Beppo Levi, 1875–1961, born in Turin, active in Piacenza, Cagliari, Parma, Bologna, and Rosario. He wrote papers in areas as distinct as algebraic geometry, set theory, integration theory, projective geometry, and number theory. Because of his Jewish origin he went into exile to Argentina in 1939.

(setting $g^+ := \max(g,0)$) then satisfy $0 \le h_n \le f_n$. From the definition of the integral it follows that

$$\sum_z (z-\varepsilon)^+ \mu(h = z, f_n > h - \varepsilon) \le \int f_n \, d\mu \; .$$

By the assumption, we have $\{f_n > h - \varepsilon\} \uparrow S$ and therefore $\mu(h = z, f_n > h - \varepsilon) \to \mu(h = z)$ thanks to σ-continuity. Consequently,

$$\sum_z (z-\varepsilon)^+ \mu(h = z) \le \lim_{n\to\infty} \int f_n \, d\mu \;,$$

and letting $\varepsilon \to 0$ we finally obtain $\sum_z z \cdot \mu(h = z) \le \lim_n \int f_n \, d\mu$. Using the definition of the integral we conclude that $\int f \, d\mu \le \lim_n \int f_n \, d\mu$, as claimed. □

A useful variant of the monotone convergence theorem is given by the following result.

Proposition 4.4 (Fatou's Lemma[3]). *Let* $f, f_1, f_2, \ldots \ge 0$ *be measurable functions satisfying the inequality* $f \le \liminf_n f_n$ *a.e. Then*

$$\int f \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu \; .$$

Proof. Setting $g_n := \inf_{m \ge n} f_m$ we get $0 \le g_1 \le g_2 \le \cdots$, $\sup_{n\ge1} g_n = \liminf_{n\to\infty} f_n$, and $g_n \le f_n$. Using the monotone convergence theorem we conclude that

$$\int f \, d\mu \le \int \liminf_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu \; .$$

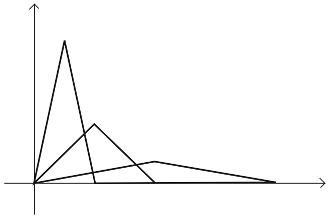
□

In Fatou's lemma, one cannot avoid the limit inferior: even if f equals the pointwise limit of f_n, in general we cannot replace the lim inf of the integrals with the lim in the assertion, as the following example shows.

[3] Pierre Fatou, 1878–1929, born in Lorient, active as astronomer at the Paris observatory. To him we owe applications of Lebesgue integration to Fourier series and complex analysis.

Example

Let (a_n) be an arbitrary sequence of positive numbers. Then $f_n := a_n n 1_{(0,1/n]}$ defines a sequence of Borel measurable mappings from $\mathbb{R}$ to $\mathbb{R}$ which converges to 0 pointwise. The Lebesgue integral $\int f_n \, d\lambda$ equals a_n and therefore may not converge. The following picture shows that the same effect can be achieved with continuous functions, too.



In order to guarantee the convergence of integrals one therefore needs additional conditions, like monotonicity in the monotone convergence theorem. In the following chapter we will encounter a different convergence criterion, namely the dominated convergence theorem.

We now compute the integral for functions taking finitely or countably infinitely many values.

Proposition 4.5. *For any measurable function* $f \geq 0$ *taking only countably many values* (*possibly including the value* ∞) *it holds that*

$$\int f \, d\mu \; = \; \sum_y y \cdot \mu(f = y) \; .$$

The sum ranges over all values y *of* f *and does not depend on the order of summation.*

Proof. Assume at first that f is elementary. If moreover h, too, is elementary and $0 \leq h \leq f$, it follows that $\mu(f = y, h = z) = \mu(\emptyset) = 0$ whenever $z > y$, therefore

$$\sum_z z \cdot \mu(h = z) \; = \; \sum_z \sum_y z \cdot \mu(h = z, f = y)$$
$$\leq \; \sum_y \sum_z y \cdot \mu(f = y, h = z) \; = \; \sum_y y \cdot \mu(f = y) \; .$$

For elementary f we thus obtain the assertion directly from the definition of the integral.

In the general case, let $y_1, y_2, \ldots$ be an arbitrary enumeration of the real values of f, and let $0 \leq z_1 \leq z_2 \ldots$ be a divergent sequence of real numbers not including any value of f. We set

$$f_n := \sum_{k=1}^{n} y_k 1_{\{f=y_k\}} + z_n \cdot 1_{\{f=\infty\}} .$$

Then $0 \leq f_1 \leq f_2 \leq \cdots$ are elementary functions with $f = \sup_n f_n$. The assertion now carries over from f_n to f with the aid of the monotone convergence theorem. □

We now employ monotone convergence in order to prove the additivity and positive homogeneity of the integral.

Proposition 4.6. *For any measurable functions* $f, g \geq 0$ *and any real numbers* $a, b \geq 0$ *we have*

$$\int (af + bg)\, d\mu \;=\; a \int f\, d\mu + b \int g\, d\mu .$$

Proof. For functions f, g having countably many values the assertion follows from Proposition 4.5, due to σ-additivity:

$$\begin{aligned}\sum_z z \cdot \mu\big(af + bg = z\big) \;&=\; \sum_z z \sum_{\substack{u,v \\ au+bv=z}} \mu\big(f = u, g = v\big) \\ &= \sum_u \sum_v (au + bv) \cdot \mu\big(f = u, g = v\big) \\ &= a \sum_u u \cdot \mu(f = u) + b \sum_v v \cdot \mu(g = v) .\end{aligned}$$

In the general case, in addition to f and g we consider the functions

$$f_n := \sum_{k=1}^{\infty} \frac{k}{2^n} \cdot 1_{\{k/2^n < f \leq (k+1)/2^n\}} + \infty \cdot 1_{\{f=\infty\}}$$

and analogously g_n. This implies $0 \leq f_1 \leq f_2 \leq \cdots$ as well as $\sup_{n \geq 1} f_n = f$. Analogous properties hold for g_n, therefore the assertion results from passing to the limit in

$$\int (af_n + bg_n)\, d\mu \;=\; a \int f_n\, d\mu + b \int g_n\, d\mu$$

by virtue of the monotone convergence theorem. □

Combining additivity and monotone convergence yields the following version of the monotone convergence theorem.

Proposition 4.7. *For arbitrary measurable functions* $f_n \geq 0$ *we have*

$$\int \sum_{n=1}^{\infty} f_n \, d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu .$$

The following proposition features an alternative formula for integrals.

Proposition 4.8. *Let* $f \geq 0$ *be measurable. Then*

$$\int f \, d\mu = \int_0^{\infty} \mu\big(f > t\big) \, dt .$$

The integral on the right hand side is to be understood as the Lebesgue integral $\int_{[0,\infty)} \mu\big(f > t\big)\, \lambda(dt)$. In the next chapter we will get on to the relation between Lebesgue- and Riemann integral.

Proof. Again we work with $f_n := \sum_{k=1}^{\infty} \frac{k}{2^n} \cdot 1_{\{k/2^n < f \leq (k+1)/2^n\}} + \infty \cdot 1_{\{f=\infty\}}$, now represented in the form

$$f_n = 2^{-n} \sum_{k=1}^{\infty} 1_{\{f > k/2^n\}} .$$

Using Proposition 4.7 we get

$$\int f_n \, d\mu = 2^{-n} \sum_{k=1}^{\infty} \mu\big(f > k/2^n\big) = \int_0^{\infty} \mu\big(f > \lceil t2^n \rceil / 2^n\big) \, dt .$$

For the left-hand side it holds that $0 \leq f_1 \leq f_2 \leq \cdots$ and $f = \sup_{n \geq 1} f_n$, while for the right-hand side we have $\lceil t2^n \rceil / 2^n \downarrow t$ and $\{f > \lceil t2^n \rceil / 2^n\} \uparrow \{f > t\}$. Passing to the limit $n \to \infty$, the assertion follows due to σ-continuity and the monotone convergence theorem. □

Already now we clearly recognize the central role played by monotone convergence within integration theory. As a method for proofs one often utilizes it in the form of the monotonicity principle, Proposition 2.8. We will illustrate this method in the next two subsections.

The Transformation Formula

Let μ be a measure on the measurable space $(S, \mathcal{A})$, let $\varphi : S \to S'$ be a $\mathcal{A}$-$\mathcal{A}'$-measurable mapping, and let

$$\mu' := \varphi(\mu)$$

be the image measure of μ under φ.

Proposition 4.9 (Transformation formula). *For any measurable* $f : S' \to \bar{\mathbb{R}}_+$ *we have*

$$\int f \, d\mu' = \int f \circ \varphi \, d\mu \, .$$

Proof. We consider

$$\mathcal{K} := \Big\{ f \geq 0 : f \text{ is measurable} \, , \int f \, d\mu' = \int f \circ \varphi \, d\mu \Big\} \, .$$

$\mathcal{K}$ satisfies the conditions (i)–(iii) of the monotonicity principle (Proposition 2.8), by virtue of the Propositions 4.3 and 4.6 and of the definition of μ. Therefore $\mathcal{K}$ includes all measurable functions $f \geq 0$. This is the assertion. □

Densities

We now introduce the notation

$$\int_A f \, d\mu := \int 1_A f \, d\mu$$

for any measurable $A \subset S$.

Definition

Let μ and ν be measures on the measurable space $(S, \mathcal{A})$. A measurable function $h \geq 0$ is called *density* of ν w.r.t. μ, if

$$\nu(A) = \int_A h \, d\mu$$

holds for all measurable sets $A \subset S$.

We then write, in short,

$$d\nu = h\,d\mu$$

or (in the style of the differential calculus)

$$h = d\nu/d\mu\ .$$

Given a measure μ and a measurable function $h \geq 0$, one may regard

$$\nu(A) \ := \int_A h\,d\mu\ , \quad A \in \mathcal{A}$$

as an equation *defining* ν. Indeed, ν is a measure on $\mathcal{A}$, the σ-additivity being a consequence of Proposition 4.7.

Proposition 4.10. *Let* $d\nu = h\,d\mu$*, and let* $f \geq 0$ *be measurable. Then*

$$\int f\,d\nu = \int fh\,d\mu\ .$$

Proof. We set

$$\mathcal{K} := \left\{ f \geq 0 : f \text{ is measurable}, \int f\,d\nu = \int fh\,d\mu \right\} .$$

By Proposition 4.6, Proposition 4.3 and the definition of a density, the assumptions (i)–(iii) of the monotonicity principle (Proposition 2.8) are satisfied. The assertion follows. □

If, in particular, $\nu = h\,d\mu$ and $\rho = k\,d\nu$, then

$$\int f\,d\rho = \int fk\,d\nu = \int fkh\,d\mu$$

resp.

$$d\rho = kh\,d\mu\ .$$

This rule is symbolically also written as

$$\frac{d\rho}{d\mu} = \frac{d\rho}{d\nu}\,\frac{d\nu}{d\mu}\ .$$

Note that, in general, densities are not uniquely determined, because if h is a density, then so is h' whenever $h' = h$ μ-a.e. holds. In the σ-finite case, however, densities are uniquely determined a.e.

Proposition 4.11. *Let* $d\nu = h\,d\mu = h'\,d\mu$ *and let* ν *be* σ*-finite. Then* $h = h'$ μ*-a.e.*

Proof. Assume first that ν is a finite measure. By Proposition 4.6,

$$\nu(h > h') + \int (h - h')^+ \,d\mu = \int\limits_{\{h>h'\}} h'\,d\mu + \int\limits_{\{h>h'\}} (h - h')^+ \,d\mu$$

$$= \int\limits_{\{h>h'\}} h\,d\mu = \nu(h > h') \ .$$

Since ν is finite, $\int (h - h')^+ \,d\mu = 0$ follows, and so $(h - h')^+ = 0$ μ-a.e. by Proposition 4.2 (iii). This means that $h \le h'$ μ-a.e. The reverse inequality follows analogously. In the σ-finite case one first considers $\int_{A_n} (h-h')^+ \,d\mu$ with $\nu(A_n) < \infty$ and then passes to the limit $n \to \infty$. □

We will get back to densities in Chap. 9 on absolute continuity.

Exercises

4.1 Let δ_x be the Dirac measure in $x \in S$. Determine $\int f\,d\delta_x$ when $f \ge 0$ is measurable.

4.2 Prove that for any measurable $f \ge 0$ and any real number $a > 0$

$$\int f^a \,d\mu = a \int_0^\infty t^{a-1} \mu(f > t)\,dt \ .$$

4.3 Let $f : \mathbb{R} \to \bar{\mathbb{R}}_+$ be a Borel measurable function satisfying $\int f\,d\lambda < \infty$, and let $a > 0$. Prove that

$$\sum_{n=1}^{\infty} n^{-a} f(nx) < \infty$$

holds for λ-almost all $x \in \mathbb{R}$.
Hint: Determine $\int f_n \,d\lambda$ for $f_n(x) := n^{-a} f(nx)$.

4.4 For measurable sets $A_1, A_2, \ldots \subset S$ we define

$$\liminf_{n\to\infty} A_n := \{x \in S : x \in A_n \text{ except for finitely many } n\} = \bigcup_{m\ge 1} \bigcap_{n\ge m} A_n \ .$$

For any measure μ, derive from the Fatou Lemma that

$$\mu\Big(\liminf_{n\to\infty} A_n\Big) \le \liminf_{n\to\infty} \mu(A_n) .$$

4.5 **Borel-Cantelli Lemma** For measurable sets $A_1, A_2, \ldots \subset S$ let

$$\limsup_{n\to\infty} A_n := \{x \in S : x \in A_n \text{ for } \infty \text{ many } n\} = \bigcap_{m\ge1} \bigcup_{n\ge m} A_n .$$

Prove that $\mu(\limsup_{n\to\infty} A_n) = 0$, assuming that $\sum_{n\ge1} \mu(A_n) < \infty$.
Hint: Consider $\int f\, d\mu$ for $f(x) := \sum_{n\ge1} 1_{A_n}(x)$, the number of those n with $x \in A_n$.

4.6 A measure μ on S is σ-finite if and only if there exists a measurable function $f \ge 0$ satisfying $\int f\, d\mu < \infty$ as well as $f(x) > 0$ for all $x \in S$. Prove this assertion.

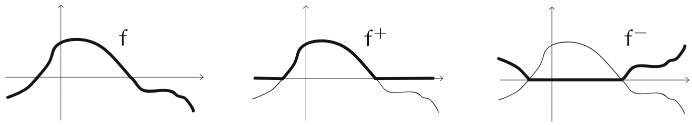
4.7 **An abstract view onto the integral** Let μ be a measure on S, and let I be a mapping, which assigns to every measurable function $f \ge 0$ a number $I(f) \ge 0$, possibly ∞, and which fulfils
(i) $f_1, f_2 \ge 0$, measurable, $c_1, c_2 \in \mathbb{R}_+ \quad \Rightarrow \quad I(c_1 f_1 + c_2 f_2) = c_1 I(f_1) + c_2 I(f_2)$,
(ii) $0 \le f_1 \le f_2 \le \cdots$ measurable $\quad \Rightarrow \quad I(\sup_n f_n) = \sup_n I(f_n)$,
(iii) $I(1_A) = \mu(A)$ for all measurable sets $A \subset S$.
Then $I(f) = \int f\, d\mu$ for all measurable $f \ge 0$.

Integrable Functions

5

Integration of measurable functions $f : S \to \bar{\mathbb{R}}$ is reduced to integration of nonnegative measurable functions as follows. We decompose f into a *positive* and a *negative* part:

$$f = f^+ - f^-, \quad \text{where} \quad f^+ := \max(f, 0) \quad \text{and} \quad f^- := \max(-f, 0)\,.$$



Definition

Let μ be a measure on S and let $f : S \to \bar{\mathbb{R}}$ be a measurable function such that $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ are not both equal to ∞. We define

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu\,.$$

In the following we will focus on functions whose integral is finite. We restrict ourselves to real-valued functions, in order to be able to add and multiply them without any restrictions.

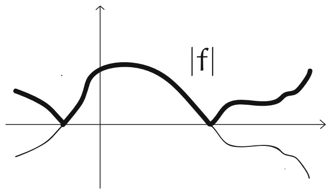
Definition

Let μ be a measure on S. A measurable function $f : S \to \mathbb{R}$ f is called *integrable*, more precisely μ-*integrable*, if $\int f^+ \, d\mu < \infty$ and $\int f^- \, d\mu < \infty$.

Since

$$|f| = f^+ + f^-\,,$$

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_5



Proposition 4.6 shows that

$$\int |f|\, d\mu = \int f^{+}\, d\mu + \int f^{-}\, d\mu\ .$$

This yields the following integrability criterion.

Proposition 5.1. *A measurable function* $f : S \to \mathbb{R}$ *is* μ*-integrable if and only if* $\int |f|\, d\mu < \infty$*, and it follows that*

$$\left| \int f\, d\mu \right| \le \int |f|\, d\mu\ .$$

Further properties of the integral arise from the results of the preceding chapter.

Proposition 5.2 (Monotonicity). *If* f, g *are integrable and satisfy* $f \le g$ *a.e., we have*

$$\int f\, d\mu \le \int g\, d\mu\ .$$

Proof. $f \le g$ a.e. implies $f^{+} + g^{-} \le f^{-} + g^{+}$ a.e. From Propositions 4.2 (i) and 4.6 we get that $\int f^{+}\, d\mu + \int g^{-}\, d\mu \le \int f^{-}\, d\mu + \int g^{+}\, d\mu$. Rearranging terms we obtain the assertion. This is permitted since all integrals are finite. □

Proposition 5.3 (Linearity). *If* f, g *are integrable and* a, b *are real numbers, then* $af + bg$*, too, is integrable, and*

$$\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu .$$

Proof. Propositions 4.2 (i) and 4.6 yield the estimate

$$\int |f+g|\,d\mu \le \int \big(|f|+|g|\big)\,d\mu = \int |f|\,d\mu + \int |g|\,d\mu < \infty\,,$$

therefore $f+g$ is integrable. From $(f+g)^+ - (f+g)^- = f+g = f^+ - f^- + g^+ - g^-$ it follows that $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$. Integrating this equation according to Proposition 4.6 and rearranging the terms one obtains $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$. The equation $\int (af)\,d\mu = a\int f\,d\mu$ is proved analogously. □

Finally, the following assertion, also called *Lebesgue's convergence theorem*, holds.

Proposition 5.4 (Dominated convergence theorem). *Let* $f_1, f_2, \ldots$ *be a sequence of measurable functions converging a.e. to a measurable function* f. *If, in addition, for some measurable function* $g \ge 0$ *with* $\int g\,d\mu < \infty$ *we have*

$$|f_n| \le g \quad \textit{a.e.}$$

for all n, *then* f_n *and* f *are integrable,* $\int |f_n - f|\,d\mu \to 0$, *and*

$$\int f_n\,d\mu \to \int f\,d\mu$$

for $n \to \infty$.

Proof. By assumption, $|f| \le g$ a.e. According to Proposition 4.2 (i), $\int |f_n|\,d\mu < \infty$ and $\int |f|\,d\mu < \infty$ follow, so f_n and f are integrable. Moreover, $2g - |f_n - f| \ge 0$ a.e., therefore Fatou's Lemma yields that

$$\int 2g\,d\mu \le \liminf_{n\to\infty} \int (2g - |f_n - f|)\,d\mu = \int 2g\,d\mu - \limsup_{n\to\infty} \int |f_n - f|\,d\mu\,.$$

Since $\int 2g\,d\mu$ is finite by assumption, it follows that $\limsup_n \int |f_n - f|\,d\mu \le 0$. Obviously we also have $0 \le \liminf_n \int |f_n - f|\,d\mu$, thus $\int |f_n - f|\,d\mu \to 0$. Since $|\int f_n\,d\mu - \int f\,d\mu| \le \int |f_n - f|\,d\mu$, we obtain the assertion. □

Additionally, we now present a generalization of the monotone convergence theorem. This result is sometimes important (e.g., in probability theory), but it will not be needed later.

A sequence $f_1, f_2, \ldots$ of $\bar{\mathbb{R}}$-valued measurable functions is called *equiintegrable*, if for every $\varepsilon > 0$ there exists an integrable function $g \ge 0$ such that

$$\sup_{n\ge 1} \int_{\{|f_n|>g\}} |f_n|\,d\mu \le \varepsilon\,.$$

Proposition 5.5. *Let the functions* $f_1, f_2, \ldots$ *be a.e. convergent to* f *and equiintegrable. Then* f_n *and* f *are integrable, and for* $n \to \infty$ *we have* $\int |f_n - f| \, d\mu \to 0$ *and*

$$\int f_n \, d\mu \to \int f \, d\mu \, .$$

Proof. Since $\{f_n\}$ is equiintegrable, so are $\{f_n^+\}$ and $\{f_n^-\}$, and they converge a.e. to f^+ resp. f^-. We therefore may assume that $f_n, f \geq 0$.

Let $\varepsilon > 0$, let $g \geq 0$ be chosen according to the equiintegrability assumption. Then we have $\int f_n \, d\mu \leq \int g \, d\mu + \varepsilon$, therefore f_n is integrable. Moreover, f is integrable, since due to $f 1_{\{f>g\}} \leq \liminf_n f_n 1_{\{f_n>g\}}$ a.e., Fatou's Lemma implies that

$$\int_{\{f>g\}} f \, d\mu \leq \liminf_{n\to\infty} \int_{\{f_n>g\}} f_n \, d\mu \leq \varepsilon \, .$$

From $|f_n - f| \leq (f_n - \min(g, f_n)) + |\min(g, f_n) - \min(g, f)| + (f - \min(g, f))$ we get that

$$\int |f_n - f| \, d\mu \leq \int_{\{f_n>g\}} f_n \, d\mu + \int |\min(g, f_n) - \min(g, f)| \, d\mu + \int_{\{f>g\}} f \, d\mu \, .$$

By the monotone convergence theorem, the middle integral on the right converges to 0, therefore

$$\limsup_{n\to\infty} \int |f_n - f| \, d\mu \leq 2\varepsilon \, .$$

Letting $\varepsilon \to 0$ we obtain $\int |f_n - f| \, d\mu \to 0$. This yields the assertion. □

In the next chapter, we will get back to the role played by equiintegrability. In particular, we will see that, conversely, $\int |f_n - f| \, d\mu \to 0$ implies the equiintegrability of $f_1, f_2, \ldots$.

Example

Let μ be a finite measure, let $\eta > 0$ and $\int |f_n|^{1+\eta}\, d\mu \le s$ for some $s < \infty$. Then for all real numbers $c > 0$ the estimate

$$\int_{\{|f_n|>c\}} |f_n|\, d\mu \le \frac{1}{c^\eta} \int_{\{|f_n|>c\}} |f_n|^{1+\eta}\, d\mu \le \frac{s}{c^\eta}$$

holds. When μ is finite, the equiintegrability of $f_1, f_2, \ldots$ results (for further elaboration see Exercise 5.5).

Two Inequalities

As a consequence of monotonicity and linearity of the integral we prove two inequalities which are based on convexity.

Proposition 5.6 (Hölder's Inequality[1]). *Let* f, g *be measurable real functions, and let* $p, q > 1$ *be conjugate real numbers, that is,* $1/p + 1/q = 1$. *If* $\int |f|^p\, d\mu < \infty$ *and* $\int |g|^q\, d\mu < \infty$, *then* fg *is integrable, and*

$$\Big| \int fg\, d\mu \Big| \le \Big(\int |f|^p\, d\mu \Big)^{1/p} \Big(\int |g|^q\, d\mu \Big)^{1/q}.$$

In the case $p = q = 2$ this is just the *Cauchy-Schwarz Inequality*[2,3]

$$\Big(\int fg\, d\mu \Big)^2 \le \int f^2\, d\mu \int g^2\, d\mu\ .$$

Proof. Since the logarithm is a concave function, for any numbers $a, b \ge 0$ we have

$$\log ab \;=\; \frac{1}{p} \log a^p + \frac{1}{q} \log b^q \;\le\; \log \Big(\frac{1}{p} a^p + \frac{1}{q} b^q \Big),$$

resp. $ab \le \frac{1}{p} a^p + \frac{1}{q} b^q$. For any $\alpha, \beta > 0$ it follows that

$$\frac{|f|}{\alpha} \cdot \frac{|g|}{\beta} \;\le\; \frac{1}{p} \frac{|f|^p}{\alpha^p} + \frac{1}{q} \frac{|g|^q}{\beta^q}.$$

[1] OTTO HÖLDER, 1859–1937, born in Stuttgart, active in Göttingen and Tübingen. He gave important contributions, in particular to group theory.

[2] AUGUSTIN-LOUIS CAUCHY, 1789–1857, born in Paris, active in Paris at the École Polytechnique and the Collège de France. He is a pioneer of real and complex analysis, all the way from the foundations to applications.

[3] HERMANN AMANDUS SCHWARZ, 1843–1921, born in Hermsdorf, Silesia, active in Zürich, Göttingen, and Berlin. His most important contributions pertain conformal mappings and the calculus of variations.

The particular choice $\alpha = (\int |f|^p \, d\mu)^{1/p}$ and $\beta = (\int |g|^q \, d\mu)^{1/q}$, in case $\alpha, \beta > 0$ results after integration in

$$\frac{1}{\alpha\beta} \int |fg| \, d\mu \;\le\; \frac{1}{p} + \frac{1}{q} \;=\; 1,$$

yielding the assertion. The case when α or β is equal to 0 has to be treated separately. If, e.g., $\int |f|^p \, d\mu = 0$, Proposition 4.2 (iii) implies that $f = 0$ a.e. and therefore $fg = 0$ a.e. as well as $\int fg \, d\mu = 0$. □

The following inequality holds for normed measures only, in general.

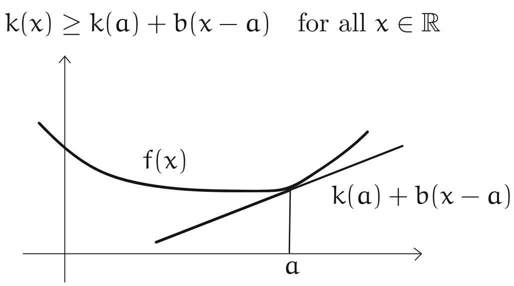
Proposition 5.7 (Jensen's Inequality[4]). *Let* μ *be a probability measure, let* f *be integrable and let the function* $k : \mathbb{R} \to \mathbb{R}$ *be convex. Then* $k \circ f$ *possesses a well-defined integral, and*

$$k\Big(\int f \, d\mu\Big) \le \int k \circ f \, d\mu \, .$$

Important special cases are given by

$$\Big|\int f \, d\mu\Big| \le \int |f| \, d\mu \, , \qquad \Big(\int f \, d\mu\Big)^2 \le \int f^2 \, d\mu \, .$$

Proof. Any convex function $k(x)$ enjoys the property of having a supporting straight line at every point a. This means that for every $a \in \mathbb{R}$ there exists a real number b such that



$$k(x) \ge k(a) + b(x - a) \quad \text{for all } x \in \mathbb{R}$$

[4] Johan Jensen, 1859–1925, born in Nakskov, active in Copenhagen for the Bell Telephone Company. He also contributed to complex analysis.

and consequently

$$k \circ f \geq k(a) + b(f - a) \ .$$

It follows that $(k \circ f)^- \leq (k(a) + b(f - a))^-$ as well as $\int (k \circ f)^- d\mu < \infty$, since f is integrable. Thus, $\int k \circ f \, d\mu$ is well-defined. In the case $\int (k \circ f)^+ \, d\mu = \infty$ the assertion now obviously holds, and so we may assume that $k \circ f$ is integrable. Due to monotonicity, linearity, and since $\mu(S) = 1$ we see that

$$\int k \circ f \, d\mu \geq k(a) + b\Big(\int f \, d\mu - a\Big) \ .$$

Setting $a = \int f \, d\mu$, the assertion follows. □

Parameter Dependent Integrals*

As an application of the dominated convergence theorem we investigate functions of the form

$$F(u) := \int f(u, x) \, \mu(dx) \ , \ u \in U \ , \quad \text{where } U \subset \mathbb{R}^d \ ,$$

concerning their continuity and differentiability.

Proposition 5.8. *Let* μ *be a measure on* S, *let* $u_0 \in U$ *and* $f : U \times S \to \mathbb{R}$ *such that*

(i) $u \mapsto f(u, x)$ *is continuous in* u_0 *for* μ*-almost all* $x \in S$,
(ii) $x \mapsto f(u, x)$ *is measurable for all* $u \in U$,
(iii) $|f(u, x)| \leq g(x)$ *for all* u, x, *for some* μ*-integrable function* $g \geq 0$.

Then F *is continuous in* u_0.

Proof. Due to (iii), $\int f(u, x) \, \mu(dx)$ is integrable for all u. The convergence of $\int f(u_n, x) \, \mu(dx)$ to $\int f(u_0, x) \, \mu(dx)$ along every sequence $u_n \to u_0$ now immediately follows from the dominated convergence theorem. □

Proposition 5.9. *Let* μ *be a measure on* S, *let* $U \subset \mathbb{R}^d$ *be open, and let* $f : U \times S \to \mathbb{R}$ *be a function having the following properties for some given* $i \in \{1, \dots, d\}$*:*

(i) $x \mapsto f(u, x)$ *is* μ*-integrable for all* u,
(ii) f *has a partial derivative w.r.t.* u_i, *and there exists a* μ*-integrable function*

$g \geq 0$ *such that for all* $u \in U$, $x \in S$

$$\left|\frac{\partial f}{\partial u_i}(u, x)\right| \leq g(x) .$$

Then F *has a partial derivative w.r.t.* u_i, $x \mapsto \frac{\partial f}{\partial u_i}(u, x)$ *is* μ*-integrable for all* $u \in U$, *and*

$$\frac{\partial F}{\partial u_i}(u) = \int \frac{\partial f}{\partial u_i}(u, x)\, \mu(dx) .$$

Proof. Since when forming partial derivatives the remaining variables are kept constant, we may assume that $d = 1$ and that U is an open interval, without loss of generality. Let $h_1, h_2, \ldots$ be a sequence converging to 0. By assumption (ii) and the mean value theorem, for all $u \in U$ we have

$$\left|\frac{f(u + h_n, x) - f(u, x)}{h_n}\right| \leq g(x) .$$

The assertion therefore results from

$$\frac{F(u + h_n) - F(u)}{h_n} = \int \frac{f(u + h_n, x) - f(u, x)}{h_n}\, \mu(dx) ,$$

passing to the limit $n \to \infty$ with the aid of the dominated convergence theorem. □

In combination with other rules of integration, one can use the preceding result to compute certain specific integrals. Examples can be found in the exercises.

Lebesgue and Riemann Integral*

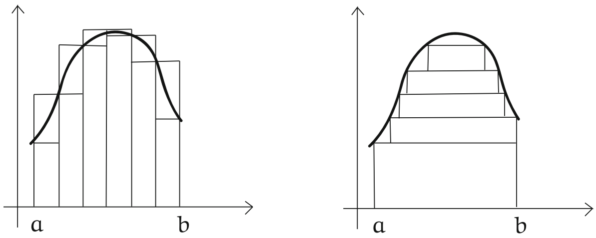
The Lebesgue integral of an integrable function f w.r.t. the Lebesgue measure we also denote by

$$\int f\, d\lambda^d = \int f(x)\, dx \quad \text{resp.} \quad \int f\, d\lambda^d = \int f(x_1, \ldots, x_d)\, dx_1 \ldots dx_d$$

and, in the case $d = 1$, also by

$$\int_{[a,b]} f\, d\lambda = \int_a^b f(x)\, dx .$$

Here we employ the notations commonly used for the Riemann integral[5] (we will recall the definition of the latter during the following proof.). Namely, it turns out that the Riemann and the Lebesgue integrals of a function f coincide if both integrals exist. The following figure visualizes the different procedures employed when integrating according to Riemann and Lebesgue, thus making it clear.



More precisely, one has the following result.

Proposition 5.10. *Let* $f : [a, b] \to \mathbb{R}$ *be a bounded* (*not necessarily measurable*) *function, and let* $C, D \subset [a, b]$ *denote its sets of continuity resp. discontinuity points. Then:*

(i) C *and* D *are Borel sets and* $f \cdot 1_C$ *is Borel measurable.*
(ii) f *is Riemann integrable if and only if* D *is a null set for the Lebesgue measure* λ, *and in this case its Riemann integral satisfies*

$$\int_a^b f(x)\,dx = \int f \cdot 1_C \,d\lambda \,.$$

For Riemann integrable functions f this does not necessarily mean that $f1_D$ is Borel measurable. However, $f1_D$ and f are measurable w.r.t. the completion of the Borel σ-algebra w.r.t. the Lebesgue measure. For this reason, the equality $\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\lambda$ makes sense.

Proof. (i) Let $a = x_0 < x_1 < \cdots < x_k = b$ be a *partition* P of the given interval having mesh size $w(P) := \max_j (x_j - x_{j-1})$. We set

$$i_j := \inf\left\{f(x) : x_{j-1} \le x \le x_j\right\}, \quad s_j := \sup\left\{f(x) : x_{j-1} \le x \le x_j\right\}$$

[5] Bernhard Riemann, 1826–1866, born in Breselenz near Hannover, active in Göttingen. His famous publications are concerned, in particular, with complex analysis, geometry, and number theory.

for $j = 1, \dots, k$, and

$$g_P := \sum_{j=1}^{k} i_j 1_{(x_{j-1},x_j]} , \quad h_P := \sum_{j=1}^{k} s_j 1_{(x_{j-1},x_j]} .$$

As is well known, the lower and upper sums of f for P are defined as

$$U_P := \sum_{j=1}^{k} i_j(x_j - x_{j-1}) = \int g_P \, d\lambda , \quad O_P := \sum_{j=1}^{k} s_j(x_j - x_{j-1}) = \int h_P \, d\lambda .$$

In the following, $P_1, P_2, \dots$ denotes a sequence of partitions such that $w(P_n)$ converges to 0 and that P_{n+1} is a refinement of P_n for all n. Then it holds that $g_{P_1} \le g_{P_2} \le \dots \le f \le \dots \le h_{P_2} \le h_{P_1}$. For the Borel measurable functions

$$g := \sup_n g_{P_n} , \quad h := \inf_n h_{P_n}$$

we get $g \le f \le h$. Since $w(P_n)$ converges to 0,

$$\{g < h\} \subset D \subset \{g < h\} \cup Q ,$$

where Q denotes the set of all partition points in $P_1, P_2, \dots$. The set Q being countable, since $\{g < h\}$ is a Borel set, then so is D, and

$$\lambda(g < h) = \lambda(D) .$$

Moreover $f \cdot 1_C = g \cdot 1_C$ holds, thus $f \cdot 1_C$ is Borel measurable, and (i) is proved.

(ii): By the dominated convergence theorem,

$$\int g \, d\lambda = \lim_n U_{P_n} , \quad \int h \, d\lambda = \lim_n O_{P_n} .$$

Since $g \le h$, we have $g = h$ a.e. if and only if $\lim_n U_{P_n} = \lim_n O_{P_n}$. In the latter case, f is called Riemann integrable (usually one considers equidistant partitions, but as we see this does not matter). Therefore, D is a null set if and only if f is Riemann integrable. Then moreover $g = f \cdot 1_C$ a.e. and

$$\int_a^b f(x) \, dx = \lim_n U_{P_n} = \int g \, d\lambda = \int f \cdot 1_C \, d\lambda ,$$

as claimed.

□

The preceding result remains valid for the d-dimensional Lebesgue measure. That the set of continuity points is Borel measurable we already know from Exercise 2.5.

The Riemann integral, while being popular in teaching, has deficiencies which render it useless for many purposes in analysis and probability theory. It lacks essential properties like the monotone convergence theorem. The Lebesgue integral mends those drawbacks.

Exercises

5.1 Let f be μ-integrable. Prove (for instance, using dominated convergence) that $n\mu(|f| \geq n) \to 0$ as $n \to \infty$.

5.2 Let $a > 1$. Prove: The measurable function $f : S \to \mathbb{R}$ is μ-integrable if and only if

$$\sum_{i=-\infty}^{\infty} a^i \mu\big(a^{i-1} \leq |f| < a^i\big) < \infty .$$

When μ is a finite measure, the condition is equivalent to

$$\sum_{n=1}^{\infty} \mu\big(|f| \geq n\big) < \infty .$$

5.3 Prove that for $n \to \infty$

$$\int_0^n (1 - x/n)^n \, dx \to \int_0^\infty e^{-x} \, dx .$$

Hint: Use $1 - t \leq e^{-t}$.

5.4 Let $f \geq 0$ be a measurable function satisfying $0 < c := \int f \, d\mu < \infty$, and let $0 < a < \infty$. Prove:

$$\lim_{n\to\infty} \int n \log(1 + (f/n)^a) \, d\mu = \begin{cases} \infty , & \text{if } a < 1, \\ c , & \text{if } a = 1, \\ 0 , & \text{if } a > 1. \end{cases}$$

Hint: Use Fatou's Lemma and the dominated convergence theorem. One has $\log(1 + x^a) \leq ax$ for all $x \geq 0, a \geq 1$.

5.5 Let μ be a finite measure and $f_1, f_2, \ldots$ be a sequence of $\bar{\mathbb{R}}$-valued measurable functions. Prove the equivalence of the following assertions:

(i) $f_1, f_2, \ldots$ is equiintegrable.

(ii) For every $\varepsilon > 0$ there exists a real number $c > 0$ such that

$$\sup_{n\geq 1} \int_{\{|f_n|>c\}} |f_n| \, d\mu \leq \varepsilon .$$

(iii) There exists a nonnegative function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\varphi(x)/x \to \infty$ for all $x \to \infty$ as well as

$$\sup_{n\geq 1} \int \varphi(|f_n|)\, d\mu < \infty .$$

One can choose φ to be convex.
Hint: Use the ansatz $\varphi(x) = \sum_{i\geq 1}(x - c_i)^+$ with $0 \leq c_i \uparrow \infty$.

5.6 Prove the equality

$$\int_0^\infty \frac{e^{-x} - e^{-ux}}{x}\, dx = \log u\,, \quad u > 0\,,$$

using differentiation.

5.7 Show that

$$F(t) := \left(\int_0^t e^{-x^2}\, dx\right)^2 + \int_0^1 \frac{e^{-t^2(x^2+1)}}{x^2+1}\, dx$$

satisfies $F(0) = \pi/4$ and $F'(t) = 0$ for all $t \geq 0$. Conclude that

$$\int_{-\infty}^\infty e^{-x^2}\, dx = \sqrt{\pi}\, .$$

5.8 Show that

$$F(t) := \int_{-\infty}^\infty e^{-x^2/2} \cos(tx)\, dx$$

satisfies $F(0) = \sqrt{2\pi}$ and $F'(t) + tF(t) = 0$. Conclude that

$$\int_{-\infty}^\infty e^{-x^2/2} \cos(tx)\, dx = \sqrt{2\pi} e^{-t^2/2}\, .$$

Convergence

6

So far we had in mind two notions of convergence for measurable functions: monotone convergence and convergence almost everywhere. Both are notions which result from convergence of the values taken by functions at fixed (but arbitrary) points of the domain. This is no longer the case for the two important notions of convergence discussed in the present chapter, convergence in the mean and convergence in measure. However, we will see that convergence almost everywhere will nevertheless come into play.

Convergence in the Mean and the Spaces $L_p(\mu)$

Definition

Let $p \geq 1$. Let μ be a measure and assume that f and $f_1, f_2, \ldots$ are real-valued measurable functions such that $\int |f|^p \, d\mu < \infty$ and $\int |f_n|^p \, d\mu < \infty$ for all n. The sequence $f_1, f_2, \ldots$ *converges in* p*-mean* (or *in* L_p*) to* f if, for $n \to \infty$,

$$\int |f_n - f|^p \, d\mu \to 0 \, .$$

In this case we write

$$f_n \xrightarrow{p} f \, .$$

It is of fundamental significance that we may interpret convergence in the mean as convergence with respect to some seminorm. We will now elaborate on this aspect. Readers who want to look up the notion of a (semi-) norm are referred to the beginning of Chap. 13.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_6

For $1 \le p < \infty$ let

$$\mathcal{L}_p(\mu) = \mathcal{L}_p(S;\mu) := \Big\{ f : S \to \mathbb{R} \ : f \text{ is measurable}, \ \int |f|^p \, d\mu < \infty \Big\} .$$

$\mathcal{L}_1(\mu)$ is just the set of all integrable functions. From the estimates

$$|f+g|^p \le (|f|+|g|)^p \le (2|f|)^p + (2|g|)^p \text{ resp.}$$
$$\int |f+g|^p \, d\mu \le 2^p \int |f|^p \, d\mu + 2^p \int |g|^p \, d\mu$$

it follows that $\mathcal{L}_p(\mu)$ is a vector space. We set

$$N_p(f) := \Big(\int |f|^p \, d\mu \Big)^{1/p} .$$

In addition, let

$$\mathcal{L}_\infty(\mu) := \big\{ f : S \to \mathbb{R} \ : f \text{ is measurable}, \ |f| \le c \text{ a.e. for some } c < \infty \big\}$$

and let

$$N_\infty(f) := \inf \{ c > 0 : |f| \le c \ \mu\text{-a.e.} \}$$

denote the *essential supremum* of $|f|$. It holds that $N_p(f) \to N_\infty(f)$ for $p \to \infty$ (exercise).

It turns out that the quantities $N_p(f)$ enjoy essential properties of a norm. Obviously,

$$N_p(af) = |a| N_p(f)$$

for every $1 \le p \le \infty$ and every real number a. It is less obvious that the triangle inequality holds.

Proposition 6.1 (Minkowski's Inequality[1]). *For any measurable functions* $f, g : S \to \mathbb{R}$ *and any* $1 \le p \le \infty$,

$$N_p(f+g) \le N_p(f) + N_p(g) .$$

[1] HERRMANN MINKOWSKI, 1864–1909, born in Kaunas, active in Bonn, Königsberg, Zürich, and Göttingen. He became famous for his contributions to number theory, convex geometry, and the theory of relativity.

Proof. For $p = 1$ the assertion follows directly from the fact that $|f+g| \le |f| + |g|$. Likewise, the case $p = \infty$ has an easy proof.

Thus, let $1 < p < \infty$. Then $1/p + 1/q = 1$ for $q := p/(p-1) > 1$. It follows that

$$\int |f+g|^p \, d\mu \le \int |f||f+g|^{p-1} \, d\mu + \int |g||f+g|^{p-1} \, d\mu$$

and, by Hölder's inequality,

$$\int |f+g|^p \, d\mu \le \Big[\Big(\int |f|^p \, d\mu\Big)^{1/p} + \Big(\int |g|^p \, d\mu\Big)^{1/p}\Big]\Big(\int |f+g|^{(p-1)q}\Big)^{1/q}$$

Since $(p-1)q = p$ and $1 - 1/q = 1/p$, the assertion follows. The particular cases $\int |f+g|^p \, d\mu = 0$ and $\int |f|^p \, d\mu = \infty$ resp. $\int |g|^p \, d\mu = \infty$ have to be considered separately, and they are trivial. □

Another important fact is that the convergence in the mean enjoys completeness.

Proposition 6.2 (Riesz[2]-Fischer[3] Theorem). *Let* $1 \le p \le \infty$*, and let* $f_1, f_2, \ldots$ *be a Cauchy sequence in* $\mathcal{L}_p(\mu)$*, that is,*

$$\lim_{m,n\to\infty} N_p(f_m - f_n) = 0 \, .$$

Then there exists an $f \in \mathcal{L}_p(\mu)$ *such that*

$$\lim_{n\to\infty} N_p(f_n - f) = 0 \, .$$

The core of the proof consists in establishing a connection to a.e. convergence by passing to a suitable subsequence.

Lemma. *Let* μ *be a measure, let* $f_1, f_2, \ldots$ *be a sequence of real-valued measurable functions satisfying*

$$\lim_{m,n\to\infty} \mu\big(|f_m - f_n| > \varepsilon\big) = 0$$

for every $\varepsilon > 0$*. Then this sequence possesses an a.e. convergent subsequence.*

[2] Frigyes Riesz, 1880–1956, born in Györ, active in Klausenburg, Szeged, and Budapest. He is mainly known due to his significant contributions to functional analysis.

[3] Ernst Fischer, 1875–1954, born in Vienna, active in Brno, Erlangen, and Köln. He also played an influential role in the development of abstract algebra. He had been forced to retire in 1938 because of his Jewish origin, but resumed teaching in Köln in 1945.

Proof. By assumption, there exists a sequence $1 \le n_1 < n_2 < \cdots$ such that, for all $m > n_k$,

$$\mu\big(|f_m - f_{n_k}| > 2^{-k}\big) \le 2^{-k} \,.$$

It follows that $\mu(|f_{n_{k+1}} - f_{n_k}| > 2^{-k}) \le 2^{-k}$. For $g := \sum_{k\ge 1} 1_{\{|f_{n_{k+1}} - f_{n_k}| > 2^{-k}\}}$, which is the number of indices k with $|f_{n_{k+1}} - f_{n_k}| > 2^{-k}$, it holds that $\int g \, d\mu = \sum_{k\ge 1} \mu(|f_{n_{k+1}} - f_{n_k}| > 2^{-k}) < \infty$. It follows that $g < \infty$ a.e., thus

$$\mu\big(|f_{n_{k+1}} - f_{n_k}| > 2^{-k} \text{ for infinitely many } k\big) = 0 \,.$$

Consequently, the series $\sum_{k\ge 1} |f_{n_{k+1}} - f_{n_k}|$ converges a.e., and therefore the sequence $f_{n_m} = f_{n_1} + \sum_{k=1}^{m-1} (f_{n_{k+1}} - f_{n_k})$ converges a.e. to a measurable function f. This completes the proof. □

Proof of the Riesz-Fischer Theorem. For $p < \infty$, due to the Markov inequality we have for every $\varepsilon > 0$

$$\mu\big(|f_m - f_n| > \varepsilon\big) \le \frac{1}{\varepsilon^p} \int |f_m - f_n|^p \, d\mu \,.$$

Thus there exist, by virtue of the assumption and the preceding lemma, a measurable function f and a subsequence $f_{n_1}, f_{n_2}, \dots$ which converges a.e. to f. By Fatou's Lemma it follows that, for all $m \ge 1$,

$$\int |f_m - f|^p \, d\mu \le \liminf_{k\to\infty} \int |f_m - f_{n_k}|^p \, d\mu \le \sup_{n\ge m} \int |f_m - f_n|^p \, d\mu \,.$$

This supremum is finite due to the assumption, thus $f - f_m$ belongs to $\mathcal{L}_p(\mu)$, and therefore so does f. Moreover, by the assumption this supremum converges to 0 as $m \to \infty$, and the assertion follows. The case $p = \infty$ is treated in a similar manner. □

The spaces $\mathcal{L}_p(\mu)$ thus enjoy properties well known in the context of Euclidean spaces. Moreover, the space $\mathbb{R}^d$ is subsumed in a natural manner. Indeed, set

$$\mu(A) := \#A \text{ for } A \subset S := \{1, \dots, d\} \,.$$

Then for $f : \{1, \dots, d\} \to \mathbb{R}$ we have

$$N_p(f) = \Big(\sum_{i=1}^{d} |f(i)|^p\Big)^{1/p} \,,$$

and in the case $p = 2$ we arrive at the usual Euclidean norm on $\mathbb{R}^d$.

One property of norms, however, does not hold: $N_p(f) = 0$ in general does not imply $f = 0$. By Proposition 4.2 one may conclude at least that $|f|^p = 0$ a.e. and thus $f = 0$ a.e. In the same manner, the limit of a sequence converging in p-mean is uniquely determined only in the a.e. sense. Therefore one cannot draw all the conclusions one is accustomed to in $\mathbb{R}^d$.

In order to remedy this deficiency one introduces new spaces. One makes use of the fact that equality a.e. is an equivalence relation, and works with equivalence classes

$$[f] := \{g : g = f \text{ a.e.}\}$$

instead of functions f. For $1 \le p \le \infty$ we set

$$L_p(\mu) := \{[f] : f \in \mathcal{L}_p(\mu)\}$$

and then for $f, g \in \mathcal{L}_p(\mu)$, $a, b \in \mathbb{R}$ define

$$a[f] + b[g] := [af + bf] ,$$
$$\|[f]\|_p := N_p(f) .$$

Obviously, all those quantities are well defined.

We may summarize our considerations as follows.

Proposition 6.3. *For $1 \le p \le \infty$, the space $L_p(\mu)$ endowed with $\|\cdot\|_p$ is a Banach space, that is, a normed vector space which is complete w.r.t. the convergence induced by the norm.*

In the case $p = 2$ we may, due to the Cauchy-Schwarz inequality, introduce a scalar product

$$([f], [g]) := \int fg \, d\mu .$$

This turns $L_2(\mu)$ into a Hilbert space, and the analogy with the Euclidean vector spaces is perfect. We will further elaborate this viewpoint in Chaps. 12 and 13.

The spaces $L_p(\mu)$ are readily designated as function spaces, and the equivalence classes are written as functions. In this manner one writes $\|f\|$ and (f, g) instead of $\|[f]\|$ and $([f], [g])$. For one thing, one often performs calculations using representatives instead of equivalence classes; for another thing, equivalence classes including a smooth function may be identified with that function.

To an equivalence class, however, one cannot in general associate a value at any single point $x \in S$ having measure 0, in contrast to what one can do for functions. Namely, for any representative one may prescribe an arbitrary value at this point.

Convergence in Measure

Convergence in measure is particularly important in probabilistic settings (there one speaks of convergence in probability).

Definition

Let μ be a measure, let $f, f_1, f_2, \dots$ be real-valued measurable functions on S. We say that the sequence $f_1, f_2, \dots$ *converges in measure* μ, or briefly *in measure to* f if

$$\lim_{n\to\infty} \mu\big(|f_n - f| > \varepsilon\big) = 0$$

for every $\varepsilon > 0$.

The limit f is a.e. uniquely determined. Namely, if $\bar{f}$ is another limit, then we have $\mu\big(|f - \bar{f}| > \varepsilon\big) = 0$, and letting $\varepsilon \to 0$ we get $\mu\big(|f - \bar{f}| > 0) = 0$.

One may motivate convergence in measure as a notion which compensates for a peculiarity of a.e. convergence. Namely, in general, it is not the case that a sequence converges a.e. if and only if every subsequence possesses an a.e. convergent subsequence. On the other hand, the following relationship holds.

Proposition 6.4. *Let* μ *be a measure, let* $f, f_1, f_2, \dots$ *be measurable real-valued functions on* S. *Consider the assertions*

(i) *The sequence* $f_1, f_2, \dots$ *converges in measure to* f,
(ii) *Each subsequence of* $f_1, f_2, \dots$ *contains a subsubsequence converging a.e. to* f.

Then (i) $\Rightarrow$ (ii). *For finite measures we even have* (i) $\Leftrightarrow$ (ii).

Proof. Assume that (i) holds. Then we may find (similarly as in the proof of the preceding lemma) for each subsequence of the natural numbers a subsubsequence $1 \le n_1 < n_2 < \cdots$ such that

$$\mu\big(|f_{n_k} - f| > 2^{-k}\big) \le 2^{-k} .$$

For $g := \sum_{k\ge 1} 1_{\{|f_{n_k} - f| > 2^{-k}\}}$ it follows that $\int g \, d\mu < \infty$ and therefore $g < \infty$ a.e., thus

$$\mu\big(|f_{n_k} - f| > 2^{-k} \text{ for infinitely many } k\big) = 0 .$$

This means that $f_{n_1}, f_{n_2}, \dots$ converges a.e. to f. Thus (ii) is proved.

Conversely, let (ii) be satisfied and let $1 \le n_1 < n_2 < \cdots$ be a subsubsequence as in (ii). For any $\varepsilon > 0$ it then follows that $1_{\{|f_{n_k} - f| > \varepsilon\}} \to 0$ a.e. as $k \to \infty$. In the case of a finite measure, the dominated convergence theorem yields

$$\mu\big(|f_{n_k} - f| > \varepsilon\big) = \int 1_{\{|f_{n_k} - f| > \varepsilon\}} \, d\mu \to 0 \, .$$

Therefore, every subsequence of the sequence $\mu\big(|f_n - f| > \varepsilon\big)$ contains a subsubsequence converging to 0. As a consequence, the whole sequence converges to 0. Thus (i) holds. □

In particular, for finite measures every sequence converging a.e. also converges in measure. The converse does not hold.

Example

Let $f_1, f_2, \ldots$ be an enumeration, in any order, of the characteristic functions $1_{I_{k,m}}$ of the intervals $I_{k,m} = [\frac{k-1}{m}, \frac{k}{m})$, with $k, m \in \mathbb{N}$ and $1 \le k \le m$, for example $f_n = 1_{I_{k,m}}$ with $n = k + m(m-1)/2$. The sequence $f_1, f_2, \ldots$ converges nowhere in the interval $[0, 1)$, but it converges in measure to 0 w.r.t. the Lebesgue measure restricted to $[0, 1)$.

Convergence in measure is superior to convergence a.e. also regarding the following completeness property.

Proposition 6.5. *Let* μ *be a measure and let* $f_1, f_2, \ldots$ *be measurable real-valued functions with the property that*

$$\lim_{m,n \to \infty} \mu\big(|f_m - f_n| > \varepsilon\big) = 0$$

for every $\varepsilon > 0$. *Then there exists a measurable function* $f : S \to \mathbb{R}$ *such that the sequence* $f_1, f_2, \ldots$ *converges in measure to* f.

Proof. By the lemma above there exists a subsequence $f_{n_1}, f_{n_2}, \ldots$ which converges a.e. to a function f. It follows that $1_{\{|f_m - f| > \varepsilon\}} \le \liminf_{k \to \infty} 1_{\{|f_m - f_{n_k}| > \varepsilon\}}$ a.e. Using Fatou's lemma we obtain for every $m \ge 1$ that

$$\mu\big(|f_m - f| > \varepsilon\big) \le \liminf_{k \to \infty} \mu\big(|f_m - f_{n_k}| > \varepsilon\big) \le \sup_{n \ge m} \mu\big(|f_m - f_n| > \varepsilon\big) \, .$$

Letting $m \to \infty$ the assertion follows in view of the assumption. □

Convergence in measure is metrizable; we will come back to this point in Exercise 6.3. We may therefore express the preceding proposition as follows: any Cauchy sequence (in such a metric) is convergent.

The Connection Between the Two Notions of Convergence*

We now relate convergence in the mean to convergence in measure. The former notion is the stronger one. More precisely, the following proposition due to F. Riesz holds.

Proposition 6.6. *Let* f *and* $f_1, f_2, \ldots$ *be elements of* $\mathcal{L}_p(\mu)$ *for some* $1 \le p < \infty$. *The following assertions are equivalent:*

(i) $f_n \xrightarrow{p} f$,
(ii) $f_1, f_2, \ldots$ *converges in measure to* f, *and* $\int |f_n|^p \, d\mu \to \int |f|^p \, d\mu$ *as* $n \to \infty$.

Proof. (i) $\Rightarrow$ (ii): The convergence in measure follows the Markov inequality

$$\mu\big(|f_n - f| \ge \varepsilon\big) \le \frac{1}{\varepsilon^p} \int |f_n - f|^p \, d\mu \, .$$

The Minkowski inequality implies that

$$\big|N_p(f_n) - N_p(f)\big| \le N_p(f_n - f) \to 0$$

and thus that $\int |f_n|^p \, d\mu \to \int |f|^p \, d\mu$.

(ii) $\Rightarrow$ (i): By Proposition 6.4, for each subsequence of the natural numbers there exists a subsubsequence $1 \le n_1 < n_2 < \cdots$ such that $f_{n_1}, f_{n_2}, \ldots$ converges a.e. to f. Fatou's lemma, applied to $2^p(|f|^p + |f_{n_k}|^p) - |f_{n_k} - f|^p \ge 0$, yields

$$2^p \int 2 \cdot |f|^p \, d\mu \le \liminf_{k\to\infty} \Big(2^p \int |f|^p \, d\mu + 2^p \int |f_{n_k}|^p \, d\mu - \int |f_{n_k} - f|^p \, d\mu\Big) \, .$$

By assumption, on the right and on the left the term $2^p \int |f|^p \, d\mu$ appears twice. By assumption it is finite, consequently

$$\limsup_{k\to\infty} \int |f_{n_k} - f|^p \, d\mu \le 0 \, .$$

Altogether each subsequence contains a subsubsequence along which $N_p(f_n - f)$ converges to 0. This is equivalent to $N_p(f_n - f) \to 0$. □

Condition (ii) of Proposition 6.6 may be further reshaped with the aid of a notion which, in a somewhat more specialized form, has already appeared in the previous chapter.

Definition

Let $p \geq 1$. A sequence $f_1, f_2, \ldots$ in $\mathcal{L}_p(\mu)$ is called *equiintegrable* (or *uniformly integrable*), more precisely *equi*-p-*integrable*, if for each $\varepsilon > 0$ there exists a measurable $g \geq 0$ satisfying $\int |g|^p \, d\mu < \infty$, such that

$$\sup_{n\geq 1} \int_{\{|f_n|>g\}} |f_n|^p \, d\mu < \varepsilon \,.$$

Replacing g by $g + |f_1| + \cdots + |f_k|$, the first k integrals under the supremum become equal to zero, for arbitrary $k \geq 1$. In this way we realize that the last requirement is equivalent to

$$\limsup_{n\to\infty} \int_{\{|f_n|>g\}} |f_n|^p \, d\mu < \varepsilon \,.$$

We will use this condition in a moment.

Proposition 6.7 (Vitali's Convergence Theorem). *Let* $f, f_1, f_2, \ldots \in \mathcal{L}_p(\mu)$ *for some* $1 \leq p < \infty$. *Then the following assertions are equivalent:*

(i) $f_n \xrightarrow{p} f$,

(ii′) $f_1, f_2, \ldots$ *are equiintegrable and converge to* f *in measure.*

Proof. (ii) ⇒ (ii′): $g := 2|f|$ belongs to $\mathcal{L}_p(\mu)$. As in the preceding proof, let $1 \leq n_1 < n_2 < \cdots$ be a subsubsequence such that $f_{n_1}, f_{n_2}, \ldots$ converges a.e. to f. Then $|f_{n_k}|^p 1_{\{|f_{n_k}|\leq g\}}$ converges a.e. to $|f|^p$, and due to the dominated convergence theorem (along subsubsequences and therefore along the whole sequence),

$$\int_{\{|f_n|\leq g\}} |f_n|^p \, d\mu \to \int |f|^p \, d\mu \,.$$

From (ii) it follows that

$$\int_{\{|f_n|>g\}} |f_n|^p \, d\mu \to 0 \,,$$

thus the equiintegrability.

(ii′) ⇒ (ii): Given $\varepsilon > 0$, choose $g \in \mathcal{L}_p(\mu)$ according to the equiintegrability condition. If we replace g by $g' := g + 2|f|$, we may conclude as before that

$$\int_{\{|f_n|\le g'\}} |f_n|^p \, d\mu \to \int |f|^p \, d\mu \, .$$

This yields

$$\limsup_{n\to\infty} \left| \int |f_n|^p \, d\mu - \int |f|^p \, d\mu \right| \le \limsup_{n\to\infty} \int_{\{|f_n|>g'\}} |f_n|^p \, d\mu < \varepsilon \, .$$

Letting $\varepsilon \to 0$ we obtain (ii). □

Exercises

6.1 Prove the Riesz-Fischer Theorem in the case $p = \infty$.

6.2 Let $f_1 \le f_2 \le \cdots$ be a sequence of measurable functions which converge in measure to a function f. Prove that the sequence converges a.e. to f.

6.3 Given measurable functions $f, g : S \to \mathbb{R}$ and a measure μ on S, let

$$d(f, g) := \inf \big\{ \varepsilon > 0 : \mu\big(|f - g| > \varepsilon\big) \le \varepsilon \big\} \, .$$

Prove: d is a pseudometric, that is, d is symmetric and satisfies the triangle inequality. d metrizes the convergence in measure, that is, $d(f_n, f) \to 0$ if and only if $f_n \to f$ in measure μ.

7 Uniqueness and Regularity of Measures

Uniqueness theorems in measure and integration theory serve to determine and identify measures. The most important of those clarifies when two measures on a σ-algebra $\mathcal{A}$ which coincide on a generator $\mathcal{E}$ of $\mathcal{A}$ are actually equal on all of $\mathcal{A}$. This is not always the case. On $\{1, 2, 3, 4\}$, for example, the system $\mathcal{E} := \big\{\{1, 2\}, \{2, 3\}\big\}$ generates the σ-algebra of all subsets, and the two probability measures μ and ν with weights $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1/4$ and $\nu_1 = \nu_3 = 1/2$, $\nu_2 = \nu_4 = 0$ coincide on $\mathcal{E}$.

For this reason, there comes into play the new condition that $\mathcal{E}$ is a $\cap$-*stable system*, meaning that

$$E, E' \in \mathcal{E} \quad \Rightarrow \quad E \cap E' \in \mathcal{E}$$

holds.

Proposition 7.1 (Uniqueness theorem for measures). *Let $\mathcal{E}$ be a $\cap$-stable generator of a σ-algebra $\mathcal{A}$ on S, and let μ, ν be two measures on $\mathcal{A}$. If*

(i) $\mu(E) = \nu(E)$ *for every* $E \in \mathcal{E}$,
(ii) $\mu(S) = \nu(S) < \infty$ *or* $\mu(E_n) = \nu(E_n) < \infty$ *for some sets* $E_1, E_2, \ldots \in \mathcal{E}$ *with* $E_n \uparrow S$,

then $\mu = \nu$.

In the case $\mu(S) = \nu(S) < \infty$ one may readily adjoin S to the generator. Thus one notices that in (ii) the second condition is more general.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_7

Example (Lebesgue Measure)

The system of all finite intervals $[a, b)$, $a, b \in \mathbb{R}^d$ is a $\cap$-stable generator of the Borel σ-algebra $\mathcal{B}^d$. It includes the intervals $[-n, n)^d$, $n \geq 1$; they have finite Lebesgue measure and their union exhausts $\mathbb{R}^d$. Therefore the Lebesgue measure λ^d is uniquely determined by its values on the intervals. This proves a part of Proposition 3.2.

Example (Borel σ-algebras)

The Borel σ-algebra $\mathcal{B}^d$ on $\mathbb{R}^d$ (or, more generally, on a metric space S) is generated by the open sets. Since S itself is open and the open sets form a $\cap$-stable system, by the theorem above a finite measure μ on $\mathcal{B}^d$ is uniquely determined by its values $\mu(O)$ for open set $O \subset S$.

Moreover, μ is uniquely determined by all integrals $\int f\, d\mu$ of bounded continuous functions f. Indeed, for any open $O \subset S$ the distance between x and O^c,

$$g(x) := d(x, O^c) = \inf\{|x - z| : z \notin O\} ,$$

is a continuous function (more precisely, we have $|g(x) - g(y)| \leq |x - y|$). The bounded continuous functions $f_n(x) := \min(1, ng(x))$ converge pointwise and monotonically towards 1_O, and the dominated convergence theorem implies that $\int f_n\, d\mu \to \mu(O)$. Therefore, μ is uniquely determined.

In order to prove the theorem we go back to the calculus of systems of sets with which we have already become acquainted in Chap. 2.

Definition

A system $\mathcal{D}$ of subsets of a nonempty set S is called a *Dynkin system,*[1] if

(i) $S \in \mathcal{D}$,
(ii) $A \in \mathcal{D} \quad \Rightarrow \quad A^c \in \mathcal{D}$,
(iii) $A_1, A_2, \ldots \in \mathcal{D} \quad \Rightarrow \quad \bigcup_{n\geq 1} A_n \in \mathcal{D}$, whenever the sets $A_1, A_2, \ldots$ are *pairwise disjoint*.

Dynkin systems are used (in contrast to σ-algebras) as a technical device only. One needs them in order to identify certain systems of sets as σ-algebras. Here it comes in handily that Dynkin systems inherit the property of $\cap$-stability from generators. This is the essence of the following fact.

Proposition 7.2. *Let $\mathcal{D}$ be a Dynkin system and $\mathcal{A}$ a σ-algebra with a $\cap$-stable generator $\mathcal{E}$. Then $\mathcal{E} \subset \mathcal{D} \subset \mathcal{A}$ implies that $\mathcal{D} = \mathcal{A}$.*

[1] EVGENII DYNKIN, born 1924 in Leningrad, active in Moscow and Cornell. He made essential contributions to Lie algebras and probability theory.

In other words, a Dynkin system generated by a $\cap$-stable system of sets is necessarily a σ-algebra.

Proof. We may assume that $\mathcal{D}$ is the smallest Dynkin system containing $\mathcal{E}$. We want to prove that whenever $\mathcal{E}$ is a $\cap$-stable system of sets, then so is $\mathcal{D}$. In order to show this, for any $D \in \mathcal{D}$ we consider the system

$$\mathcal{D}_D := \{A \in \mathcal{D} : A \cap D \in \mathcal{D}\} .$$

$\mathcal{D}_D$, too, is a Dynkin system: Properties (i) are (iii) are obviously satisfied. Moreover, for any $A \in \mathcal{D}_D$ the disjoint union $(A \cap D) \cup D^c$ and hence its complement $A^c \cap D$ belongs to $\mathcal{D}$. Thus property (ii) holds.

Now let $E \in \mathcal{E}$. We get $\mathcal{E} \subset \mathcal{D}_E$, since $\mathcal{E}$ is $\cap$-stable by assumption. The minimality of $\mathcal{D}$ yields $\mathcal{D}_E = \mathcal{D}$, in other words: $D \cap E \in \mathcal{D}$ for every $D \in \mathcal{D}$, $E \in \mathcal{E}$. This means that $\mathcal{E} \subset \mathcal{D}_D$ holds for every $D \in \mathcal{D}$. Again, the minimality of $\mathcal{D}$ yields the equality $\mathcal{D}_D = \mathcal{D}$, this time for every $D \in \mathcal{D}$. By definition, this equality means that $\mathcal{D}$ is $\cap$-stable, as claimed.

Now, in $\mathcal{D}$ we may convert every countable union into a disjoint union, according to the scheme

$$\bigcup_{n \geq 1} A_n = A_1 \cup \bigcup_{n \geq 2} (A_n \cap A_1^c \cap \cdots \cap A_{n-1}^c) .$$

Therefore, $\mathcal{D}$ is a σ-algebra. Since $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{E}$, the assertion follows. □

Proof of the uniqueness theorem. Let $E_n \in \mathcal{E}$ such that $\mu(E_n) = \nu(E_n) < \infty$ and $E_n \uparrow S$. By the properties of measures,

$$\mathcal{D}_n := \big\{A \in \mathcal{A} : \mu(A \cap E_n) = \nu(A \cap E_n)\big\}$$

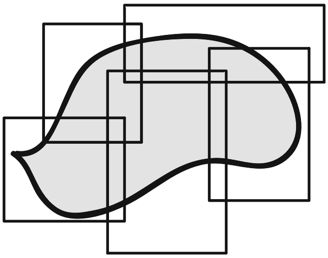
is a Dynkin system. Since $\mathcal{E}$ is $\cap$-stable, it follows that $\mathcal{E} \subset \mathcal{D}_n \subset \mathcal{A}$ and moreover, due to the preceding proposition, that $\mathcal{D}_n = \mathcal{A}$. Therefore $\mu(A \cap E_n) = \nu(A \cap E_n)$ holds for each $A \in \mathcal{A}$. Passing to the limit $n \to \infty$ we obtain the assertion. □

Regularity*

We now treat situation where the values $\mu(E)$ of a measure on a generator $\mathcal{E}$ and its other values $\mu(A)$ are related in a more explicit manner. To this purpose we form the expression

$$\mu^*(A) := \inf\Big\{\sum_{m \geq 1} \mu(E_m) : E_1, E_2, \ldots \in \mathcal{E}, A \subset \bigcup_{m \geq 1} E_m\Big\} , \quad A \subset S ,$$

which is determined solely by the restriction of μ on $\mathcal{E}$. One thus considers finite or countably infinite coverings of A with elements from the generator such that the sum of their measures is as small as possible.



From the properties of measures (monotonicity and sub-σ-additivity) it follows that

$$\mu(A) \leq \mu^*(A)$$

for every $A \in \mathcal{A}$. Moreover,

$$\mu(E) = \mu^*(E)$$

for every $E \in \mathcal{E}$, because E is covered by itself. When is it possible to conclude $\mu(A) = \mu^*(A)$ for other sets $A \in \mathcal{A}$, too? This question leads us, following Carathéodory,[2] to the following definition (it is a bit more general than usual: we do not restrict ourselves to Borel σ-algebras).

Definition

Let μ be a measure on a σ-algebra $\mathcal{A}$, let $\mathcal{E}$ be a generator of $\mathcal{A}$. μ is called *outer regular* (with respect to $\mathcal{E}$), if

$$\mu(A) = \inf\Big\{\sum_{m\geq 1}\mu(E_m) : E_1, E_2, \ldots \in \mathcal{E},\ A \subset \bigcup_{m\geq 1} E_m\Big\}$$

holds for every $A \in \mathcal{A}$.

Some generators are immediately ruled out at this point, for example the generator of the Borel σ-algebra in $\mathbb{R}$ consisting of the intervals $(-\infty, x] \subset \mathbb{R}$, which cannot be used to cover perfectly arbitrary Borel sets. However, even when using more suitable generators not every measure is outer regular.

---

[2] CONSTANTIN CARATHÉODORY, 1873–1950, born in Berlin, active at several German universities, in Athens, and finally from 1924 in Munich. He made essential contributions to measure and integration theory, the calculus of variations, complex analysis, and the axiomatic treatment of thermodynamics. During the period 1920–1922 he acted as founding rector of the university at Smyrna.

Example

The counting measure $\mu(B) := \#B$ on the Borel σ-algebra in $\mathbb{R}$, as well as the σ-finite measure $\mu(B) := \#B \cap \mathbb{Q}$, do not possess outer regularity with respect to the generator formed by the open sets.

Usually one deals with measures which are outer regular with respect to a clearly specified generator. This is true at least for measures constructed by Carathéodory's method. We will discuss this in Chap. 11.

When outer regularity is present, we may supplement the uniqueness theorem for measures with the following comparison result, which is sometimes useful.

Proposition 7.3 (Comparison Theorem). *Let* μ *and* ν *be measures on a* σ-*algebra* $\mathcal{A}$ *with generator* $\mathcal{E}$. *If*

$$\nu(E) \le \mu(E)$$

for every $E \in \mathcal{E}$, *and if* μ *is outer regular w.r.t.* $\mathcal{E}$, *then* $\nu \le \mu$.

Proof. Let $A \in \mathcal{A}$ and $A \subset \bigcup_{m \ge 1} E_m$ with $E_m \in \mathcal{E}$. By the properties of measures and the assumption,

$$\nu(A) \le \sum_{m \ge 1} \nu(E_m) \le \sum_{m \ge 1} \mu(E_m) .$$

Taking the infimum over all coverings of A we conclude from the outer regularity that $\nu(A) \le \mu(A)$, as asserted. □

In particular, an outer regular measure is maximal among all measures that coincide on $\mathcal{E}$.

We now pursue the question how we may read off outer regularity from the generator.

Proposition 7.4. *Let* $\mathcal{E}$ *be a* $\cap$-*stable generator of a* σ-*algebra* $\mathcal{A}$ *on* S *satisfying* $\emptyset \in \mathcal{E}$. *Let* μ *be a measure on* $\mathcal{A}$, *and assume that there exist sets* $E_1, E_2, \ldots \in \mathcal{E}$ *such that* $E_n \uparrow S$ *and* $\mu(E_n) < \infty$ *for every* $n \ge 1$. *If*

$$\mu(E' \setminus E) = \mu^*(E' \setminus E) \quad \textit{for every } E, E' \in \mathcal{E} \textit{ with } E \subset E' ,$$

then μ *is outer regular w.r.t.* $\mathcal{E}$.

We will present the proof in Chap. 11.

Example (Semirings, Outer Regularity of the Lebesgue Measure)

A $\cap$-stable system $\mathcal{E}$ of sets with $\emptyset \in \mathcal{E}$ is called a *semiring*, if for every $E, E' \in \mathcal{E}$ with $E \subset E'$ there exist disjoint sets $E_1, E_2, \ldots \in \mathcal{E}$ such that $E' \setminus E = \bigcup_{m\geq 1} E_m$. In this case we have $\sum_{m\geq 1} \mu(E_m) = \mu(E' \setminus E)$ and consequently

$$\mu^*(E' \setminus E) = \mu(E' \setminus E) .$$

If, in addition, S can be exhausted with elements E_n, $n \geq 1$, from $\mathcal{E}$ with finite measure, then the assumptions of the preceding proposition are satisfied and μ is outer regular.

In particular this shows that the d-dimensional Lebesgue measure is outer regular with respect to the generator $\mathcal{E}$ of the Borel σ-algebra $\mathcal{B}^d$ consisting of all d-dimensional intervals

$$E = [a, b) , \quad a, b \in \mathbb{R}^d .$$

Obviously, $\mathcal{E}$ is a semiring, and moreover $\mathbb{R}^d = \bigcup_{m\geq 1}[-m, m)^d$ and $\lambda^d\big([-m, m)^d\big) < \infty$.

We now discuss the important case where the generator consists of all open subsets of a metric space. Proposition 7.4 yields the following result.

Proposition 7.5. *Let μ be a measure on the Borel σ-algebra of a metric space S, and assume that there exist open sets $E_1, E_2, \ldots \subset S$ satisfying $E_n \uparrow S$ and $\mu(E_n) < \infty$ for each $n \geq 1$. Then μ is outer regular. More precisely, for every Borel set B we have*

$$\mu(B) = \inf\big\{\mu(O) : O \supset B, O \textit{ is open}\big\}$$

as well as

$$\mu(B) = \sup\big\{\mu(A) : A \subset B, A \textit{ is closed}\big\} .$$

Proof. First, let μ be finite. We check the assumptions of the preceding proposition: The open sets form a $\cap$-stable system of sets which includes the empty set.

Furthermore, let $O \subset O'$ be open. Then $A := O^c$ is closed, so for any null sequence $\varepsilon_1 > \varepsilon_2 > \cdots > 0$ of real numbers we have

$$O^c = \bigcap_{n=1}^{\infty} A^{\varepsilon_n}$$

where $A^\varepsilon := \{x \in S : d(x, y) < \varepsilon \text{ for some } y \in A\}$ (the open ε-neighbourhood of A in the metric d). Since μ is finite, using σ-continuity we get that $\mu(O' \setminus O) = \lim_{n\to\infty} \mu(A^{\varepsilon_n} \cap O')$. Moreover, $O' \setminus O$ is covered by the open sets $A^\varepsilon \cap O'$, for each $\varepsilon > 0$. In conclusion,

$$\mu^*(O' \setminus O) = \mu(O' \setminus O) ,$$

therefore by the preceding proposition μ is outer regular. The first assertion now follows since a union of open sets is open. Passing to complements reveals that the second assertion is equivalent to the first.

More generally, let now $E_1 \subset E_2 \subset \cdots$ be open sets of finite measure exhausting S. As we have just proved, the proposition applies to the finite measures $\mu(\cdot \cap E_m)$. For any Borel set $B \subset S$ and any $\varepsilon > 0$ there now exist closed sets A_m and open sets O_m satisfying $A_m \subset B \subset O_m$ and $\mu(O_m \cap E_m) < \mu(A_m \cap E_m) + \varepsilon 2^{-m}$. Setting $A := \bigcup_{m\geq 1} A_m$ and $O := \bigcup_{m\geq 1} O_m \cap E_m$ we get $A \subset B \subset O$ and $\mu(O) < \mu(A) + \varepsilon$. Moreover, by σ-continuity, $\mu\big(\bigcup_{m=1}^n A_m\big) \to \mu(A)$ as $n \to \infty$. Since $\bigcup_{m=1}^n A_m$ is closed and O is open, the assertion follows. □

For measures on Borel σ-algebras one further expands the notion of regularity.

Definition

A measure μ on a Borel σ-algebra is called *outer regular*, if for every Borel set B,

$$\mu(B) = \inf\big\{\mu(O) : O \supset B, O \text{ is open}\big\} .$$

μ is called *inner regular*, if for every Borel set B,

$$\mu(B) = \sup\big\{\mu(K) : K \subset B, K \text{ is compact}\big\} .$$

If both properties hold, μ is called *regular*.

Proposition 7.6. *Let μ be a measure on a metric space S which satisfies the assumptions of the preceding proposition. Suppose that S is a K_σ-set, that is, there exist compact sets $K_n \subset S$, $n \geq 1$ with $K_n \uparrow S$. Then μ is regular.*

Proof. Due to σ-continuity, the assumption implies that $\mu(A \cap K_n) \to \mu(A)$ as $n \to \infty$. When A is closed, the sets $A \cap K_n$ are compact. The assertion thus follows from the preceding proposition. □

Example (Regularity of the Lebesgue measure)

Setting $K_n = [-n, n]^d$, the measure λ^d obviously satisfies the assumptions of the proposition.

For the further development of measure theory on topological spaces, *Radon measures* play a prominent role. Radon measures are those regular measures on Borel σ-algebras, which are locally finite, that is, for which every $x \in S$ possesses an open neighbourhood of finite measure. We do not dwell further on this issue.

Density of the Continuous Functions*

As an application of the regularity of the Lebesgue measure obtained just above, we prove that the continuous functions are dense in the spaces $\mathcal{L}_p(\lambda^d)$. Recall that the *support* of a continuous function $g : \mathbb{R}^d \to \mathbb{R}$ is defined as the topological closure of the set $\{x \in \mathbb{R}^d : g(x) \neq 0\}$.

Proposition 7.7. *Let* $f \in \mathcal{L}_p(\lambda^d)$*, where* $1 \le p < \infty$*. Then for every* $\varepsilon > 0$ *there exists a continuous function* $g : \mathbb{R}^d \to \mathbb{R}$ *with compact support such that*

$$\int \left|f(x) - g(x)\right|^p dx < \varepsilon .$$

Proof. We first consider the case $f = 1_B$, where $B \subset \mathbb{R}^d$ is any Borel set with $\lambda^d(B) < \infty$. Since the Lebesgue measure is regular, for every $\varepsilon > 0$ there exist a compact set K and an open set O satisfying $K \subset B \subset O$ and $\lambda^d(O) < \lambda^d(K) + \varepsilon$. Due to compactness, there exists a $\delta > 0$ such that $|x - y| \ge \delta$ for all $x \in K, y \notin O$. The function

$$g(x) := (1 - \delta^{-1} d(x, K))^+ , \quad \text{where } d(x, K) := \inf\{|x - y| : y \in K\}$$

is continuous. Its support is contained in the closed δ-neighbourhood of K and therefore compact. g takes values between 0 and 1, on K the value 1 and on O^c the value 0. This implies $|1_B - g|^p \le 1_{O \setminus K} = 1_O - 1_K$ and

$$\int |1_B - g|^p \, d\lambda^d \le \mu(O) - \mu(K) < \varepsilon ,$$

which proves the assertion for $f = 1_B$.

When $f \in \mathcal{L}_p(\lambda^d)$ is arbitrary, for every $\varepsilon > 0$ there exist natural numbers m, n such that $\int |f - f'|^p \, d\lambda^d < \varepsilon$ for $f' = \sum_{k=-n}^{n} \frac{k}{m} 1_{\{k/m \le f < (k+1)/m\}}$. The summands can be approximated as described above by continuous functions with compact support. In this manner, the assertion follows for general f (here the Minkowski inequality helps). □

Exercises

7.1 Let μ be a finite measure on $S_1 \times S_2$ (endowed with the product σ-algebra). Let μ_1 and μ_2 be the two image measures of μ under the projection mappings π_1 und π_2. Prove by means of an example that μ is not uniquely determined by μ_1 and μ_2 (even though π_1 and π_2 generate the product σ-algebra).

7.2 Let S be a finite set and $\mathcal{D}$ the system of all subsets consisting of an even number of elements. When is $\mathcal{D}$ a Dynkin system? Is $\mathcal{D}$ then a σ-algebra?

7.3 Let $\mathcal{D}$ be a Dynkin system and $A, A' \in \mathcal{D}$ where $A' \subset A$. Prove that $A \setminus A'$ also belongs to $\mathcal{D}$.
Hint: Consider $(A \setminus A')^c$.

7.4 Let $\mathcal{M}$ be the smallest set of functions f from a metric space S to $\mathbb{R}$ with the properties
(i) $f_n \in \mathcal{M}, f_n \to f$ pointwise $\Rightarrow$ $f \in \mathcal{M}$
(ii) $\mathcal{M}$ includes all continuous functions.
Prove that $\mathcal{M}$ equals the set of all Borel measurable functions.
Hint: In order to show that $\mathcal{M}$ is a vector space, for any given $g \in \mathcal{M}$, $a, b \in \mathbb{R}$ consider the set $\mathcal{M}_{g,a,b} := \{f \in \mathcal{M} : af + bg \in \mathcal{M}\}$, at first for continuous g, and then for arbitrary $g \in \mathcal{M}$. Prove in addition that $\mathcal{D} := \{B \in \mathcal{B} : 1_B \in \mathcal{M}\}$ is a Dynkin system which contains the open sets.

7.5 Show that for any Lebesgue integrable function $f : \mathbb{R} \to \mathbb{R}$ we have $\int \big|f(x+t)-f(x)\big|\, dx \to 0$ as $t \to 0$.
Hint: First consider the case of a continuous function with compact support.

7.6 **Steinhaus' Theorem** Let $B \subset \mathbb{R}$ be a Borel set with $\lambda(B) > 0$. Prove that $B-B := \{x-y : x, y \in B\}$ contains an interval $(-\delta, \delta)$ for some $\delta > 0$.
Hint: Conclude from the preceding exercise that $\lambda\big(B \cap (B+t)\big) \to \lambda(B)$ as $t \to 0$.

Multiple Integrals and Product Measures

8

It is not particularly surprising that one may integrate measurable functions multiply with respect to different variables. But it did irritate mathematicians like Cauchy that the result may depend on the sequential arrangement of the integrals. When computing derivatives this is usually not the case.

Only with the advent of Lebesgue's integration theory it turned out that in integration, too, the result usually does not depend on the order in which the integrals are taken. This is the content of Fubini's theorem, a core result of this chapter. It has theoretical significance, but is also relevant when computing specific integrals. Some important examples will be found in the text, others in the exercises.

Multiple integrals have many applications. We will construct product measures and discuss the convolution and smoothing of functions. Finally, we will address a more general situation: the integration of kernels.

Double Integrals

Multiple integration rests on the following fact.

Lemma. *Let* $(S', \mathcal{A}')$, $(S'', \mathcal{A}'')$ *be measurable spaces,* ν *a* σ*-finite measure on* $\mathcal{A}''$*, and* $f : S' \times S'' \to \bar{\mathbb{R}}_+$ *a nonnegative* $\mathcal{A}' \otimes \mathcal{A}''$*-*$\bar{\mathcal{B}}$*-measurable function. Then the following assertions hold.*

(i) *The mapping* $y \mapsto f(x, y)$ *is* $\mathcal{A}''$*-*$\bar{\mathcal{B}}$*-measurable for each* $x \in S'$*. Consequently, the integral* $\int f(x, y)\, \nu(dy)$ *is well-defined for each* $x \in S'$.
(ii) *The mapping* $x \mapsto \int f(x, y)\, \nu(dy)$ *is nonnegative and* $\mathcal{A}'$*-*$\bar{\mathcal{B}}$*-measurable.*

Proof. We restrict ourselves to the case of a finite measure ν (the σ-finite case is a consequence). We consider the system $\mathcal{D}$ of those sets $A \in \mathcal{A}' \otimes \mathcal{A}''$ for which the

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_8

function $f = 1_A$ satisfies the assertions (i) and (ii). By the properties of measurable mappings and Proposition 4.7, $\mathcal{D}$ includes all unions of disjoint sets $A_1, A_2, \ldots$ from $\mathcal{D}$. Since ν is assumed to be finite, $\mathcal{D}$ also includes the complement A^c for each A in $\mathcal{D}$. Finally, $S' \times S''$ belongs to $\mathcal{D}$, therefore $\mathcal{D}$ is a Dynkin system.

Moreover, $A' \times A'' \in \mathcal{D}$ for all $A' \in \mathcal{A}'$, $A'' \in \mathcal{A}''$, as one concludes from the equality $1_{A' \times A''}(x, y) = 1_{A'}(x) 1_{A''}(y)$. Because those product sets form a $\cap$-stable generator of the product σ-algebra, it follows from Proposition 7.2 that $\mathcal{D}$ coincides with the product σ-algebra.

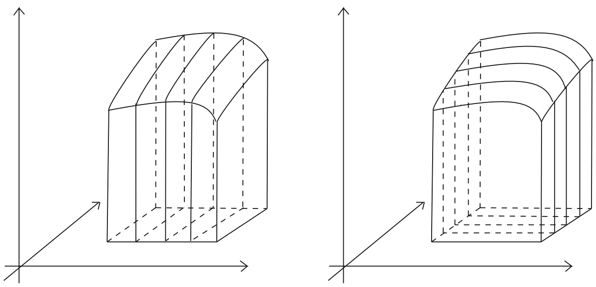
Let $\mathcal{K}$ be the system of all nonnegative $\mathcal{A}' \otimes \mathcal{A}''$-$\bar{\mathcal{B}}$-measurable functions $f : S' \times S'' \to \bar{\mathbb{R}}$ satisfying assertions (i) and (ii). By what we just proved and due to the properties of measurable functions and integrals, $\mathcal{K}$ satisfies the conditions of the monotonicity principle (Proposition 2.8). Therefore, $\mathcal{K}$ includes all nonnegative $\mathcal{A}' \otimes \mathcal{A}''$-$\bar{\mathcal{B}}$-measurable functions $f : S' \times S'' \to \bar{\mathbb{R}}$. The assertion is proved. □

For σ-finite measures μ and ν and nonnegative measurable functions f the double integral

$$\int \Big(\int f(x, y)\, \nu(dy) \Big) \mu(dx) = \int_{S'} \Big(\int_{S''} f(x, y)\, \nu(dy) \Big) \mu(dx) \tag{8.1}$$

is thus well-defined, as is the double integral with the order of integrations interchanged.

It is a fundamental fact that the order in which the integrations are performed does not matter.



Proposition 8.1 (Fubini). *For σ-finite measures μ and ν on σ-algebras $\mathcal{A}'$ and $\mathcal{A}''$ and nonnegative measurable functions* $f : S' \times S'' \to \bar{\mathbb{R}}_+$ *we have*

$$\int \Big(\int f(x, y)\, \nu(dy) \Big) \mu(dx) = \int \Big(\int f(x, y)\, \mu(dx) \Big) \nu(dy)\, .$$

Proof. Once more we restrict ourselves to the case of finite measures. We consider the system $\mathcal{D}$ of those sets $A \in \mathcal{A}' \otimes \mathcal{A}''$ for which our assertion holds with 1_A in place of f. Using the properties of integrals we again conclude that $\mathcal{D}$ is a Dynkin system. For $f(x, y) := 1_{A'\times A''}(x, y) = 1_{A'}(x)1_{A''}(y)$ both integrals are equal to $\mu(A')\nu(A'')$, thus $A' \times A'' \in \mathcal{D}$, and $\mathcal{D}$ again coincides with the product σ-algebra.

We form the system $\mathcal{K}$ of all nonnegative $\mathcal{A}' \otimes \mathcal{A}''$-$\bar{\mathcal{B}}$-measurable functions $f : S' \times S'' \to \bar{\mathbb{R}}$ for which the asserted equation holds. By what we just proved and due to the properties of integrals, $\mathcal{K}$ satisfies the conditions in Proposition 2.8, and our proposition follows. □

Example

We have

$$\int_0^\infty \Big(\int_0^\infty e^{-(1+x^2)y^2} y\, dy\Big) dx = \int_0^\infty \Big(\int_0^\infty e^{-(1+x^2)z} \frac{1}{2}\, dz\Big) dx$$
$$= \frac{1}{2}\int_0^\infty \frac{1}{1+x^2}\, dx = \frac{1}{2}\big[\arctan x\big]_0^\infty = \frac{\pi}{4}$$

and

$$\int_0^\infty \Big(\int_0^\infty e^{-(1+x^2)y^2} y\, dx\Big) dy = \int_0^\infty e^{-y^2}\Big(\int_0^\infty e^{-(xy)^2} y\, dx\Big) dy$$
$$= \int_0^\infty e^{-y^2}\Big(\int_0^\infty e^{-z^2}\, dz\Big) dy = \Big(\int_0^\infty e^{-z^2}\, dz\Big)^2 .$$

By Fubini's Theorem the two expressions are equal, and we obtain the important formula

$$\int_{-\infty}^\infty e^{-z^2}\, dz = \sqrt{\pi}\, .$$

This argument is due to Laplace,[1] the formula itself had already been obtained earlier by Euler.[2]

So far we have written the double integrals with parentheses, in order to be precise. In the following, according to common usage we will omit them.

We now introduce double integrals for measurable real-valued functions f(x, y) which may take negative values, too. As in the case of single integrals this is not

[1] Pierre-Simon Laplace, 1749–1827, born in Beaumont-en-Auge, active in Paris at the École Militaire and École Polytechnique. His main research areas were celestial mechanics and probability theory.

[2] Leonard Euler, 1707–1783, born in Basel, active in St. Petersburg and Berlin. He shaped mathematics far beyond his century.

always possible; here the additional assumption is

$$\iint \left|f(x, y)\right| \mu(dx)\nu(dy) < \infty ,$$

where according to the theorem of Fubini the order of integration is immaterial. Whenever this holds in addition to measurability as above, f is called integrable.

But even under this assumption a small cliff has to be circumvented: It is still possible that $\int f^+(x, y)\, \nu(dy)$ as well as $\int f^-(x, y)\, \nu(dy)$ may attain the value ∞ for some values of x, so that we may not be able to form the integral $\int f(x, y)\, \nu(dy)$ as we did before. However the set of those x's is a μ-null set. More precisely, the following lemma holds.

Lemma. *Let* $f : S' \times S'' \to \mathbb{R}$ *be measurable, le* μ *and* ν *be* σ*-finite measures and* $\iint \left|f(x, y)\right| \mu(dx)\nu(dy) < \infty$*. Then there exists a measurable function* $\hat{f} : S' \times S'' \to \mathbb{R}$ *with the following properties:*

(i) *It holds* $\hat{f} = f$ *a.e., that is,* $\hat{f}(x, \cdot) = f(x, \cdot)$ ν*-a.e. for* μ*-almost all* $x \in S'$*,*
(ii) $y \mapsto \hat{f}(x, y)$ *is* ν*-integrable for all* $x \in S'$*,*
(iii) $x \mapsto \int \hat{f}(x, y)\, \nu(dy)$ *is* μ*-integrable.*

Proof. Let A' be the set of those $x \in S'$ for which $\int \left|f(x, y)\right| \nu(dy) < \infty$. We set $\hat{f}(x, y) := f(x, y)1_{A'}(x)$. By assumption, $\iint \left|f(x, y)\right| \nu(dy)\mu(dx) < \infty$. By Proposition 4.2 (iv), $\mu\big((A')^c\big) = 0$ follows. This yields assertion (i). (ii) holds due to the choice of A'. From $\iint \left|f(x, y)\right| \nu(dy)\mu(dx) = \iint \left|\hat{f}(x, y)\right| \nu(dy)\mu(dx)$ we get

$$\int \left| \int \hat{f}(x, y)\, \nu(dy) \right| \mu(dx) \le \iint \left|\hat{f}(x, y)\right| \nu(dy)\mu(dx) < \infty ,$$

and thus (iii) holds. □

In particular, the lemma implies that $\int f(x, y)\, \nu(dy)$ exists for μ-almost all $x \in S'$. For $\hat{f}$, due to (ii) and (iii) we can form the double integral $\iint \hat{f}(x, y)\, \nu(dy)\mu(dx)$ (with the order of integration as indicated!). If $\tilde{f}$ is another measurable function with the properties stated in the lemma, then applying Proposition 4.2 (ii) twice it follows according to property (i) that

$$\iint \tilde{f}(x, y)\, \nu(dy)\mu(dx) = \iint \hat{f}(x, y)\, \nu(dy)\mu(dx).$$

Thus if we assume that $\iint |f(x,y)|\, \mu(dx)\nu(dy) < \infty$, the integral

$$\iint f(x,y)\, \nu(dy)\mu(dx) := \iint \hat{f}(x,y)\, \nu(dy)\mu(dx)$$

is well-defined. Its value is finite. In an analogous manner one obtains the double integral in reverse order.

The properties of the double integral again result from decomposing f into its positive and negative parts. One has $\int \hat{f}(x,y)\, \nu(dy) = \int \hat{f}^+(x,y)\, \nu(dy) - \int \hat{f}^-(x,y)\, \nu(dy)$. If f is integrable, those integrals viewed as functions of x are μ-integrable. The linearity of the integral yields

$$\iint \hat{f}(x,y)\, \nu(dy)\mu(dx) = \iint \hat{f}^+(x,y)\, \nu(dy)\mu(dx) - \iint \hat{f}^-(x,y)\, \nu(dy)\mu(dx)\,.$$

In addition, $\hat{f}^+$ and $\hat{f}^-$ are a.e. equal to f^+ and f^-, therefore we obtain – in this case not by the definitions, but via the detour of integrating $\hat{f}$ – the equation

$$\iint f(x,y)\, \nu(dy)\mu(dx) = \iint f^+(x,y)\, \nu(dy)\mu(dx) - \iint f^-(x,y)\, \nu(dy)\mu(dx)\,.$$

In the right-hand side we may apply the standard integration rules and thus arrive at the properties of double integrals. In particular, we obtain a second version of Fubini's Theorem.

Proposition 8.2 (Fubini). *Let the measurable real-valued function* $f : S' \times S'' \to \bar{\mathbb{R}}$ *and the σ-finite measures* μ, ν *satisfy* $\iint |f(x,y)|\, \nu(dy)\mu(dx) < \infty$. *Then*

$$\iint f(x,y)\, \nu(dy)\mu(dx) = \iint f(x,y)\, \mu(dx)\nu(dy)\,.$$

Example (Reordering of absolutely convergent series)

For any doubly-indexed sequence $f(m,n)$ of real numbers with $\sum_{m\ge 1}\sum_{n\ge 1} |f(m,n)| < \infty$ (absolute convergence) it holds that

$$\sum_{m\ge 1}\sum_{n\ge 1} f(m,n) = \sum_{n\ge 1}\sum_{m\ge 1} f(m,n)\,.$$

We may view this as a particular case of Fubini's Theorem, applied to the σ-finite counting measures $\mu(A) = \nu(A) = \#A$, $A \subset \mathbb{N}$. The requirement of absolute convergence cannot just be omitted, as the example $f(m,m) = 1$, $f(m,m+1) = -1$ and $f(m,n) = 0$ otherwise shows.

Here we have

$$\sum_{m\geq 1}\sum_{n\geq 1} f(m,n) = 1 \neq 0 = \sum_{n\geq 1}\sum_{m\geq 1} f(m,n) .$$

As in this example, in many specific instances one may choose $\hat{f}(x,y) = f(x,y)$. The above-mentioned problem concerning the existence of integrals does not arise.

Multiple integrals may easily be reduced to double integrals. Details are left to the reader.

Product Measures

Double integrals give rise to new measures on the product σ-algebra.

Proposition 8.3. *Let μ and ν be σ-finite measures on the σ-algebras $\mathcal{A}'$ and $\mathcal{A}''$. Then*

$$\pi(A) := \iint 1_A(x,y)\,\nu(dy)\mu(dx) , \quad A \in \mathcal{A}' \otimes \mathcal{A}'' ,$$

defines a measure π on the product σ-algebra. We have

$$\int f\,d\pi = \iint f(x,y)\,\nu(dy)\mu(dx)$$

for every measurable function $f \geq 0$.

Proof. Obviously $\pi(\varnothing) = 0$ holds, and the σ-additivity results from applying Proposition 4.7 twice. In order to prove the second assertion we consider

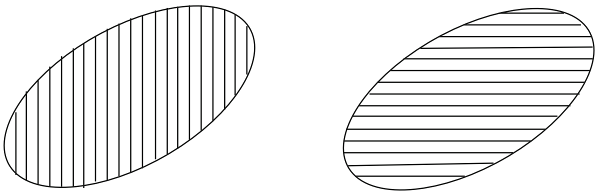
$$\mathcal{K} := \left\{ f \geq 0 : \int f\,d\pi = \iint f(x,y)\,\nu(dy)\mu(dx) \right\} .$$

Due to the definition of π, $\mathcal{K}$ includes all elements A of the product σ-algebra. By virtue of the rules of integration, the other two conditions of the monotonicity principle (Proposition 2.8) are satisfied, too. Therefore $\mathcal{K}$ includes all nonnegative measurable functions, and the assertion follows. □

According to the exposition in the foregoing section it also holds that

$$\int f\,d\pi = \iint f(x,y)\,\mu(dx)\nu(dy) ,$$

that is one can reverse the order of integration.



One calls π the *product measure* of μ and ν, and writes

$$\pi = \mu \otimes \nu \quad \text{or} \quad \pi(dx, dy) = \mu(dx) \otimes \nu(dy) .$$

$f(x, y)$ is $\mu \otimes \nu$-integrable if and only if

$$\int |f| \, d(\mu \otimes \nu) = \iint |f(x, y)| \, \mu(dx)\nu(dy) < \infty .$$

In this case, integrals w.r.t. the product measure can be reduced to double integrals, the order of integration being arbitrary. This fact, too, is called the *Fubini's Theorem.*

► **Remark** A set $A \in \mathcal{A}' \otimes \mathcal{A}''$ is a $\mu \otimes \nu$-null set if and only if the double integral $\iint 1_A(x, y) \; \nu(dy)\mu(dx) = \int \nu(A_x) \, \mu(dx)$ equals 0, employing the "cross section" $A_x := \{y \in S'' : (x, y) \in A\}$. In other words, A is a $\mu \otimes \nu$-null set if and only if A_x is a ν-null set for μ-almost all $x \in S'$. This is in complete accordance with the a.e. notion used in item (i) of the foregoing lemma.

The following proposition illuminates why one speaks of "product measures".

Proposition 8.4. *Let* μ *and* ν *be* σ*-finite measures. Then*

$$(\mu \otimes \nu)(A' \times A'') = \mu(A') \cdot \nu(A'')$$

for all $A' \in \mathcal{A}'$, $A'' \in \mathcal{A}''$. *These product formula determines* $\mu \otimes \nu$ *uniquely.*

Proof. The product formula results from the double integral of the function $1_{A' \times A''}(x, y) = 1_{A'}(x) 1_{A''}(y)$. The other assertion follows from the uniqueness theorem for measures, as μ and ν are assumed to be σ-finite, and the measurable sets of the form $A' \times A''$ form a $\cap$-stable generator of the product σ-algebra. □

Example (Lebesgue Measure)

We recall that the Borel σ-algebras in $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ (thus $d = d_1 + d_2$) satisfy $\mathcal{B}^d = \mathcal{B}^{d_1} \otimes \mathcal{B}^{d_2}$. The Cartesian product $[a_1, b_1) \times [a_2, b_2) \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ of semi-open intervals is again a semi-open interval, and

$$\lambda^d\big([a_1, b_1) \times [a_2, b_2)\big) = \lambda^{d_1}\big([a_1, b_1)\big) \cdot \lambda^{d_2}\big([a_2, b_2)\big) .$$

Consequently, $\lambda^d\big([a, b)\big) = \lambda^{d_1} \otimes \lambda^{d_2}\big([a, b)\big)$ holds for each $[a, b) \subset \mathbb{R}^d$. Since these semi-open intervals form a ∩-stable generator of the Borel σ-algebra, the uniqueness theorem for measures implies that

$$\lambda^d = \lambda^{d_1} \otimes \lambda^{d_2} .$$

In this manner Lebesgue integrals may be reduced to multiple integrals, and we obtain the formula

$$\int f\, d\lambda^d = \int \cdots \int f(x_1, \ldots, x_d)\, dx_1 \ldots dx_d .$$

Example (Volume of the d-dimensional unit ball)

We want to determine the volume

$$v_d := \lambda^d(B_1)$$

of the unit ball $B_1 := \{x \in \mathbb{R}^d : |x| \le 1\}$ in $\mathbb{R}^d$ through reduction to the Γ-function

$$\Gamma(t) := \int_0^\infty e^{-z} z^{t-1}\, dz , \quad t > 0 .$$

For this purpose we consider the image measure $\mu = \varphi(\lambda^d)$ of the Lebesgue measure under the mapping $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ given by $\varphi(x) = |x|^2$. Due to the transformation formula for integrals in Chap. 4 we have

$$\int e^{-y}\, \mu(dy) = \int e^{-|x|^2}\, \lambda^d(dx) .$$

Let us calculate the two integrals. Since λ^d is a product measure, a multiple application of Fubini's Theorem yields that

$$\int e^{-|x|^2}\, \lambda^d(dx) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-x_1^2} \cdots e^{-x_d^2}\, dx_1 \ldots dx_d = \Big(\int_{-\infty}^{\infty} e^{-u^2}\, du \Big)^d .$$

For the other integral we use the formula $\mu\big([0, z]\big) = \lambda^d\big(z^{1/2} B_1\big) = z^{d/2} v_d$, $z > 0$. From Fubini's Theorem and since $e^{-y} = \int_y^\infty e^{-z}\, dz$ it follows that

$$\begin{aligned} \int e^{-y}\, \mu(dy) &= \int_0^\infty \int_0^\infty e^{-z} 1_{\{y \le z\}}\, dz\, \mu(dy) \\ &= \int_0^\infty e^{-z} \int_0^\infty 1_{\{y \le z\}} \mu(dy)\, dz = \int_0^\infty e^{-z} \mu\big([0, z]\big)\, dz = v_d \Gamma\Big(\frac{d}{2} + 1\Big) . \end{aligned}$$

The comparison of the two integrals is revealing already in the cases $d = 1$ and 2: As is well known, we have $v_2 = \pi$ (area of the unit circle), and $\Gamma(2) = 1$ (partial integration). It follows that $(\int_{-\infty}^{\infty} e^{-u^2}\,du)^2 = \pi$, a formula which we have derived already. From $v_1 = 2$ it then follows that $\sqrt{\pi} = 2\Gamma(3/2)$.

In conclusion

$$v_d = \frac{\pi^{d/2}}{\Gamma\big(d/2+1\big)}\ .$$

We may evaluate the Γ-function inductively using the formula $\Gamma(t+1) = t\Gamma(t)$ (partial integration) as well as the values just obtained, namely $\Gamma(2) = 1$ and $\Gamma(3/2) = \sqrt{\pi}/2$ resp. $\Gamma(1/2) = \sqrt{\pi}$. Details are left to the reader.

Convolution and Smoothing*

We apply multiple integration to a particular situation. Let $g, h : \mathbb{R}^d \to \mathbb{R}$ be Lebesgue integrable functions, thus

$$\int \big|g(x)\big|\,dx\ ,\ \int \big|h(x)\big|\,dx < \infty\ .$$

The function $f(x,y) := g(x-y)h(y)$ is Borel measurable on $\mathbb{R}^{2d}$, and since we have $\int |g(x-y)|\,dx = \int |g(x)|\,dx$ we get

$$\iint \big|g(x-y)h(y)\big|\,dxdy = \int \big|g(x)\big|\,dx \int \big|h(y)\big|\,dy < \infty\ . \qquad (*)$$

In the section above on double integrals we have seen that the *convolution integral*

$$\int g(x-y)h(y)\,dy$$

then exists for almost all x, resp. it defines a Lebesgue integrable function uniquely for a.e. $x \in \mathbb{R}^d$. Moreover, the convolution integral remains unchanged if g or h are modified on Lebesgue null sets. Therefore it is natural to understand g, h and their convolution integral as equivalence classes of measurable functions, as elements of $L_1(\lambda^d)$. We thus give the following definition:

Definition

Let $g, h \in L_1(\lambda^d)$. Their *convolution* $g * h \in L_1(\lambda^d)$ is defined as

$$g * h(x) := \int g(x-y)h(y)\,dy\ .$$

By (∗) we have

$$\|g * h\|_1 \le \|g\|_1 \|h\|_1 \ .$$

Convolutions arise in various situations.

Example

Let $a > 0$, and let $k : [0, \infty) \to \mathbb{R}$ be continuous. The solution of the inhomogeneous linear differential equation

$$f'(x) = af(x) + k(x) \ , \quad x \ge 0 \ ,$$

together with the boundary condition $f(0) = 0$ is given by

$$f(x) = \int_0^x k(y)e^{-a(x-y)}\, dy = \int g(x-y)h(y)\, dy$$

where $g(x) := e^{-ax}$, $h(x) := k(x)$ for $x \ge 0$ and $g(x) = h(x) := 0$ for $x < 0$. One can verify this directly by differentiation.

Convolution is important also because of its nice algebraic properties. When we substitute $y \mapsto x - y$, the convolution integral changes into $\int g(y)h(x-y)\, dy$, and consequently

$$g * h = h * g \ .$$

Moreover,

$$(g * h) * k = g * (h * k) \ , \ g * (h + k) = g * h + g * k \ .$$

We leave the proof as an exercise for the reader.
We now explain how functions can be smoothed by convolution, and thus show that the smooth functions are dense in $L_p(\lambda^d)$. For every $\delta > 0$ we choose a so-called "mollifier" $\kappa_\delta : \mathbb{R}^d \to \mathbb{R}$ with the following properties:

(a) κ_δ is nonnegative and differentiable of infinite order,
(b) $\kappa_\delta(x) = 0$ for $|x| \ge \delta$,
(c) $\int \kappa_\delta(x)\, dx = 1$.

We may take, for example, $\kappa_\delta(x) := \delta^{-d}\kappa(\delta^{-1}x)$ with

$$\kappa(x) := \begin{cases} c \exp\big(-(1-|x|^2)^{-1}\big) \ , & \text{if } |x| < 1 \ , \\ 0 \ , & \text{if } |x| \ge 1 \ , \end{cases}$$

where $c > 0$ is a suitably chosen normalization constant.

For any measurable function $f : \mathbb{R}^d \to \mathbb{R}$ we define, whenever $\int |f|^p \, d\lambda^d < \infty$ for some $p \geq 1$, the functions

$$f_\delta := f * k_\delta ,$$

or

$$f_\delta(x) = \int f(y) k_\delta(x-y) \, dy .$$

The integral exists in the case $p = 1$ because κ_δ is bounded, and in the case $p > 1$ as a consequence of Hölder's inequality. Applying Proposition 5.9 on the differentiation of integrals we see that f_δ has derivatives of any order.

Proposition 8.5 (Smoothing Theorem). *Let* $1 \leq p < \infty$. *Then for* $f \in \mathcal{L}_p(\lambda^d)$ *we have*

$$\|f - f * \kappa_\delta\|_p \to 0$$

as $\delta \to 0$.

Proof. First we prove the result for continuous g with compact support. As is known, g then is uniformly continuous. For any given $\varepsilon > 0$ it holds for sufficiently small $\delta > 0$ that $|g(x) - g(x-y)| \leq \varepsilon$ for each $|y| \leq \delta$. Consequently,

$$\big|g(x) - g * \kappa_\delta(x)\big| \leq \int \big|g(x) - g(x-y)\big| k_\delta(y) \, dy \leq \varepsilon .$$

Therefore, $g * \kappa_\delta$ converges uniformly to g. Moreover, $g(x)$ and thus $g * \kappa_\delta(x)$ are nonzero in a bounded domain only. It follows that $\|g * \kappa_\delta - g\|_p \to 0$ as $\delta \to 0$, as one checks with the aid of the dominated convergence theorem.

The transition from a continuous g with compact support to arbitary $f \in \mathcal{L}_p(\lambda^p)$ we achieve with an estimate. Since $\int k_\delta \, d\lambda^d = 1$, Jensen's inequality shows that (note that $t \mapsto |t|^p$ is convex for $p \geq 1$)

$$\begin{aligned} \|f * \kappa_\delta\|_p^p &= \int \Big| \int f(x-y) \kappa_\delta(y) \, dy \Big|^p dx \leq \iint \big|f(x-y)\big|^p \kappa_\delta(y) \, dy dx \\ &= \int \Big(\int \big|f(x-y)\big|^p dx \Big) \kappa_\delta(y) \, dy = \|f\|_p^p . \end{aligned}$$

According to Proposition 7.7, for any given $\varepsilon > 0$ we choose a continuous g with compact support such that $\|f - g\|_p < \varepsilon$. It follows that

$$\|f - f * \kappa_\delta\|_p \leq \|f - g\|_p + \|g - g * \kappa_\delta\|_p + \|(g - f) * \kappa_\delta\|_p \leq 2\varepsilon + \|g - g * \kappa_\delta\|_p .$$

Letting $\delta \to 0$ we get $\limsup_{\delta\to 0} \|f - f * \kappa_\delta\|_p \le 2\varepsilon$, and letting $\varepsilon \to 0$ we obtain the assertion. □

Kernels*

We discuss a generalization which is important in probability theory: within the double integral $\int(\int f(x,y)\,\nu(dy))\mu(dx)$ one allows the measure ν to depend on x. In order that the outer integral be defined, one needs a regularity assumption.

Definition

Let $(S', \mathcal{A}')$, $(S'', \mathcal{A}'')$ be measurable spaces. A family

$$\nu = \big(\nu(x, dy)\big)_{x\in S'}$$

of finite measures $\nu(x, dy)$ on $\mathcal{A}''$ is called a *kernel* of $(S', \mathcal{A}')$ to $(S'', \mathcal{A}'')$, if for every $A'' \in \mathcal{A}''$ the function

$$x \mapsto \nu(x, A'')$$

is $\mathcal{A}'$-$\mathcal{B}^1$-measurable.

Lemma. *Let* ν *be a kernel of* $(S', \mathcal{A}')$ *to* $(S'', \mathcal{A}'')$ *and let* $f : S' \times S'' \to \bar{\mathbb{R}}_+$ *be a nonnegative* $\mathcal{A}' \otimes \mathcal{A}''$-$\bar{\mathcal{B}}$*-measurable function. Then the function*

$$x \mapsto \int f(x, y)\,\nu(x, dy)$$

is $\mathcal{A}'$-$\bar{\mathcal{B}}$*-measurable.*

Proof. As before, we consider the system $\mathcal{D}$ of sets $A \in \mathcal{A}' \otimes \mathcal{A}''$ for which the function $f = 1_A$ satisfies the assertion. By the properties of measurable functions and Proposition 4.7, $\mathcal{D}$ includes the union of disjoint sequences $A_1, A_2, \dots$ as well as complements of sets in $\mathcal{D}$. Finally, $S' \times S''$ is included in $\mathcal{D}$ by the measurability properties of kernels, therefore $\mathcal{D}$ is a Dynkin system.

Moreover, we have $A' \times A'' \in \mathcal{D}$ for all $A' \in \mathcal{A}'$, $A'' \in \mathcal{A}''$, as one sees from the equation $\int 1_{A'\times A''}(x, y)\,\nu_x(dy) = 1_{A'}(x)\nu(x, A'')$. Because these product sets form a $\cap$-stable generator of the product σ-algebra, we conclude from Proposition 7.2 that $\mathcal{D}$ coincides with the product σ-algebra.

The assertion now follows in the same manner as in the proof of the lemma at the beginning of this chapter. □

Thus one again may form double integrals. For reasons of clarity we use the notation

$$\int \mu(dx) \int \nu(x, dy)\, f(x, y)\ .$$

Once more,

$$A \mapsto \int \mu(dx) \int \nu(x, dy)\, 1_A(x, y)$$

defines a measure on the product σ-algebra, denoted as

$$\mu \otimes \nu \quad \text{resp.} \quad \mu(dx) \otimes \nu(x, dy)\ .$$

It is an interesting question which measures can be obtained by this procedure, that is, under which conditions can a given measure π on the product σ-algebra be expressed as $\pi = \mu \otimes \nu$ for some measure μ and some kernel ν. One speaks of *disintegration* of the measure π. On Borel σ-algebras this is possible under rather general conditions. We do not dwell further on this subject.

Exercises

8.1 Prove and comment on the following observation, due to Cauchy: the double integrals

$$\int_{(0,1)} \int_{(0,1)} \frac{x^2 - y^2}{(x^2 + y^2)^2}\, dxdy\ , \quad \int_{(0,1)} \int_{(0,1)} \frac{x^2 - y^2}{(x^2 + y^2)^2}\, dydx$$

are well-defined and different from each other.
Hint: $(x^2 - y^2)(x^2 + y^2)^{-2} = \partial^2 \arctan(x/y)/\partial x \partial y$.

8.2 Let μ be the counting measure on $\mathbb{R}$, that is, $\mu(B) := \#B$ for Borel sets $B \subset \mathbb{R}$, and let D be the diagonal in $\mathbb{R}^2$, $D = \{(x, y) \in \mathbb{R}^2 : x = y\}$. Prove and comment:

$$\iint \mathbf{1}_D(x, y)\, \lambda(dx)\mu(dy) \neq \iint \mathbf{1}_D(x, y)\, \mu(dy)\lambda(dx)\ .$$

8.3 Let $\nu_1(dx) = h_1(x)\, \mu_1(dx)$, $\nu_2(dy) = h_2(y)\, \mu_2(dy)$. What is the density of $\nu_1 \otimes \nu_2$ w.r.t. $\mu_1 \otimes \mu_2$?

8.4 **Integrals "measure the area below a function"** Let $f : S \to \bar{\mathbb{R}}^+$ be measurable. Prove the formula

$$\int f\, d\mu = \mu \otimes \lambda(A_f) = \int_0^\infty \mu(f > t)\, dt\ ,$$

where $A_f = \{(x, t) \in S \times \mathbb{R} : 0 \le t < f(x)\}$.
Hint: It holds that $f(x) = \int 1_{\{0 \le t < f(x)\}}\, dt$. Concerning the measurability of A_f compare Exercise 2.7.

8.5 We want to derive the formula (improper Riemann integral)

$$\int_0^\infty \frac{\sin x}{x}\,dx = \frac{\pi}{2}\,.$$

First, prove that $\int_0^a \int_0^\infty x \exp(-xy)\,dydx < \infty$ for every $0 \le a < \infty$. Conclude that

$$\int_0^a \frac{\sin x}{x}\,dx = \int_0^\infty \int_0^a \sin x\, e^{-xy}\,dxdy\,.$$

Compute the inner integral, most simply as the imaginary part of $\int_0^a \exp((i-y)x)\,dx$, and pass to the limit $a \to \infty$.

8.6 In the same manner prove that for $a > 0$ one has

$$\int_0^\infty \frac{e^{-x} - e^{-ax}}{x}\,dx = \log a\,.$$

8.7 **The Beta function** The Beta function is defined as

$$B(x, y) := \int_0^1 s^{x-1}(1-s)^{y-1}\,ds\,, \quad x, y > 0\,.$$

We want to express it with the aid of the Gamma function $\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}\,dt$, $x > 0$. Prove that

$$\Gamma(x+y)B(x, y) = \int_0^\infty \Big(\int_0^t u^{x-1}(t-u)^{y-1}\,du\Big)e^{-t}\,dt\,.$$

Using Fubini's Theorem and a shift of variable conclude that

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}\,.$$

8.8 Prove that $(g * h) * k = g * (h * k)$.

Absolute Continuity

9

In this chapter we discuss under which circumstances measures and functions have densities.

In the first situation considered, two measures μ and ν are given on some σ-algebra, and we ask for conditions under which a measureable function h exists such that $d\nu = h\, d\mu$, that is,

$$\nu(A) = \int_A h\, d\mu$$

holds for all measurable sets A. In the second situation, a function $f : [a, b] \to \mathbb{R}$ is given and one asks for the existence of a Borel measurable function $h : [a, b] \to \mathbb{R}$ such that

$$f(x) = \int_a^x h(z)\, dz$$

holds for all $x \in [a, b]$.

The two questions are related. This becomes clear when one chooses for μ the Lebesgue measure, restricted to the interval [a, b], and for ν another measure on the Borel sets in [a, b]. Setting $f(x) := \nu\big([a, x]\big)$ and $A := [a, x]$, the first equation turns into the second.

Therefore, it is possible to treat both problem statements simultaneously. However, we want to consider two different methods; for measures a "global" exhaustion procedure, for functions a "local" method which is more involved, but provides a connection to differentiation and to the fundamental theorem of calculus.

For measures, one may easily formulate a necessary condition for the existence of a density. One obviously has to require that

$$\mu(A) = 0 \;\Rightarrow\; \nu(A) = 0\,.$$

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_9

We will show that for σ-finite measures this condition is also sufficient. As we will see in the exercises, the following seemingly stronger condition is equivalent:

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall A \in \mathcal{A}\ :\ \mu(A) \leq \delta\ \Rightarrow\ \nu(A) \leq \varepsilon\ .$$

For functions, we will consider an analogous condition.

Absolute Continuity and Singularity of Measures

This section is about the following pair of complementary notions.

Definition

Let μ and ν be two measures on a σ-algebra $\mathcal{A}$.

(i) ν is called *absolutely continuous* w.r.t. μ, written as

$$\nu \ll \mu\ ,$$

if, for each $A \in \mathcal{A}$, $\mu(A) = 0$ implies that $\nu(A) = 0$. If μ and ν are both absolutely continuous w.r.t. each other, μ and ν are called *equivalent*.

(ii) μ and ν are called mutually *singular*, written as

$$\mu \perp \nu\ ,$$

if there exists an $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$.

Absolute continuity can be characterized as follows.

Proposition 9.1 (Radon[1]-Nikodym[2] Theorem). *Let μ and ν be σ-finite measures on a σ-algebra $\mathcal{A}$. Then the following assertions are equivalent:*

(i) $\nu \ll \mu$,

(ii) $d\nu = h\, d\mu$ *for some measurable function* $h : S \to \mathbb{R}_+$.

The density h *then is* μ*-a.e. finite and* μ*-a.e. unique.*

[1] Johann Radon, 1887–1956, born in Tetschen, active a.o. in Hamburg, Breslau, and Vienna. His working areas were measure and integration theory, functional analysis, calculus of variations, and differential geometry.

[2] Otton Nikodým, 1887–1974, born in Zablotow, active in Kraków, Warsaw and at Kenyon College, Ohio. He worked on measure theory and functional analysis.

One may after all drop the requirement that ν is σ-finite. Concerning μ this is not the case (compare Exercise 9.2).

From the various proofs available we choose a lucid classical approach. It uses a result which moreover is of independent interest.

Proposition 9.2 (Hahn Decomposition[3]). *Let ν and ρ be finite measures on a σ-algebra $\mathcal{A}$. Then there is a measurable set $A_\le$, its complement being $A_\ge := S \setminus A_\le$, such that*

$$\nu(A) \le \rho(A) \quad \textit{for all } A \subset A_\le ,$$
$$\nu(A) \ge \rho(A) \quad \textit{for all } A \subset A_\ge .$$

Proof. We set $\delta(A) := \nu(A) - \rho(A)$ for $A \in \mathcal{A}$. Then like a measure, δ satisfies $\delta(\emptyset) = 0$ and is σ-additive, but $\delta(A)$ may be negative. For later purposes we allow $\delta(A)$ to assume the value ∞, but not the value $-\infty$.

(i) A measurable set $N \subset S$ we term *negative* if $\delta(A) \le 0$ for all $A \subset N$. We want to construct $A_\le$ as the largest possible negative set. It fits our purpose that whenever $N_1, N_2, \ldots \subset S$ are negative, then so is $\bigcup_{k\ge1} N_k$. Namely, for $A \subset \bigcup_{k\ge1} N_k$ the set $A_k := A \cap N_k \cap N_1^c \cap \cdots \cap N_{k-1}^c$ is a subset of N_k, thus $\delta(A_k) \le 0$ and $\delta(A) = \sum_{k\ge1} \delta(A_k) \le 0$ follows.

(ii) First, in the case $\delta(S) < \infty$ we construct a negative subset $N \subset S$ satisfying $\delta(N^c) \ge 0$. We obtain N by removing successively certain disjoint measurable sets B_k, $k \ge 1$, with $\delta(B_k) \ge 0$, for which $\delta(B_k)$ is sufficiently large. We set $B_1 := \emptyset$. Having chosen $B_1, \ldots, B_k$ we let s_k be the supremum of all values $\delta(A)$, taken over those measurable sets A which are disjoint with $B_1, \ldots, B_k$. It holds $s_k \ge \delta(\emptyset) = 0$. Now we choose the set B_{k+1} as disjoint to $B_1, \ldots, B_k$, and such that $\delta(B_{k+1}) \ge s_k/2$ in case $s_k < \infty$, in particular $B_{k+1} = \emptyset$ in case $s_k = 0$, and $\delta(B_{k+1}) \ge 1$ in case $s_k = \infty$.

Let now $N := S \setminus \bigcup_{k\ge1} B_k$. Then we have $\delta(N^c) = \sum_{k\ge1} \delta(B_k)$, thus $\delta(N^c) \ge 0$. From $\delta(N) + \delta(N^c) = \delta(S) < \infty$ it follows that $\delta(N^c) < \infty$. This implies $\delta(B_k) \to 0$ and thus $s_k \to 0$. If $A \subset N$, then A is disjoint with $B_1, \ldots, B_k$, and therefore $\delta(A) \le s_k$. Passing to the limit $k \to \infty$ we obtain $\delta(A) \le 0$. Thus N is negative.

(iii) More generally, we claim that if $S' \subset S$ is measurable and $\delta(S') < \infty$, then there exists a negative set $N' \subset S'$ such that $\delta(S' \setminus N') \ge 0$ and therefore $\delta(N') \le \delta(S')$. This follows from (ii) when we consider the restriction δ' of δ to the measurable subsets of S'.

[3] Hans Hahn, 1879–1934, born in Vienna, active in Chernovitz, Bonn, and Vienna. He made essential contributions to functional analysis, measure theory, and real analysis. He played a leading role in the Vienna Circle, a group of positivist philosophers and scientists.

(iv) Let now $\alpha := \inf\{\delta(A) : A \in \mathcal{A}\}$, thus $\alpha \le 0$. Let $S_k \subset S$, $k \ge 1$, be measurable subsets satisfying $\delta(S_k) < \infty$ and $\delta(S_k) \to \alpha$. According to (iii), there exist negative sets $N_k \subset S_k$ such that $\delta(N_k) \le \delta(S_k)$. It follows that $\delta(N_k) \to \alpha$. We set $A_\le := \bigcup_{k\ge 1} N_k$. By (i) the set $A_\le$ is negative. Therefore we have $\delta(A_\le) = \delta(A_\le \setminus N_k) + \delta(N_k) \le \delta(N_k)$ for all k and thus $\delta(A_\le) = \alpha$. It follows that $\alpha > -\infty$. We now finish the proof as follows:

Let $A \subset A_\le$. Since $A_\le$ is negative, we have $\delta(A) \le 0$, resp. $\nu(A) \le \rho(A)$. This is one part of the assertion. On the other hand, let $A \subset S \setminus A_\le$. Then we have $\delta(A) = \delta(A \cup A_\le) - \delta(A_\le) \ge \alpha - \alpha = 0$. This is the other part of the assertion.

□

Proof of the Radon-Nikodym Theorem. The implication (ii) ⇒ (i) is obviously true. To prove (i) ⇒ (ii) let us first assume that μ and ν are finite. We consider the set

$$\mathcal{F} := \left\{ f \ge 0 \ : \ \int_A f \, d\mu \le \nu(A) \text{ for all } A \in \mathcal{A} \right\}$$

of measurable functions and set

$$\beta := \sup_{f \in \mathcal{F}} \int f \, d\mu \ .$$

As ν is finite, we have $\beta \le \nu(S) < \infty$. We want to obtain the sought-after density h as an element of $\mathcal{F}$ satisfying

$$\int h \, d\mu = \beta \ .$$

To this end we claim that $\max(f, f') \in \mathcal{F}$ whenever $f, f' \in \mathcal{F}$. Indeed, the latter gives

$$\int_A \max(f, f') \, d\mu = \int_{A \cap \{f \ge f'\}} f \, d\mu + \int_{A \cap \{f < f'\}} f' \, d\mu$$
$$\le \nu\big(A \cap \{f \ge f'\}\big) + \nu\big(A \cap \{f < f'\}\big) = \nu(A) \ .$$

If $f_1, f_2, \dots$ are elements of $\mathcal{F}$ satisfying $\int f_n \, d\mu \to \beta$, we may assume without loss of generality that $0 \le f_1 \le f_2 \le \cdots$, otherwise replace f_n with $\max(f_1, \dots, f_n)$. For $h := \sup_n f_n$ using monotone convergence we obtain that $h \in \mathcal{F}$ and $\int h \, d\mu = \beta$.

For any $A' \in \mathcal{A}$ we therefore have that $\int_{A'} h \, d\mu \le \nu(A')$. To prove the reverse inequality we consider, for any given $\varepsilon > 0$, the finite measure ρ having the density $d\rho = (h + \varepsilon 1_{A'}) \, d\mu$, and additionally, according to the foregoing proposition, the Hahn decomposition $A_\le, A_\ge$ for ν and ρ. On $A_\le$, ν is dominated by ρ, and thus we get the estimate

$$\nu(A' \cap A_{\leq}) \leq \rho(A' \cap A_{\leq}) \leq \rho(A') = \int_{A'} h \, d\mu + \varepsilon\mu(A') \, .$$

On $A_{\geq}$, ρ remains below ν. Therefore, $g := h + \varepsilon 1_{A' \cap A_{\geq}}$ belongs to $\mathcal{F}$, since for any measurable A we have

$$\int_A g \, d\mu = \rho(A \cap A_{\geq}) + \int_{A \cap A_{\leq}} h \, d\mu \leq \nu(A \cap A_{\geq}) + \nu(A \cap A_{\leq}) = \nu(A) \, .$$

From $\int g \, d\mu = \beta + \varepsilon\mu(A' \cap A_{\geq})$ it follows that $\mu(A' \cap A_{\geq}) = 0$ and, since $\nu \ll \mu$, we get $\nu(A' \cap A_{\geq}) = 0$. All in all, it follows that

$$\nu(A') \leq \int_{A'} h \, d\mu + \varepsilon\mu(A') \, ,$$

and letting $\varepsilon \to 0$ we obtain the desired inequality.

Thus, $d\nu = h \, d\mu$. In particular, $\nu(h = \infty) = \infty \cdot \mu(h = \infty)$. Since ν is finite, we also get $h < \infty \mu$-a.e. The μ-a.e. uniqueness was established above.

These results easily carry over to σ-finite measures if we exhaust S by a sequence of sets of finite measure. □

The Radon-Nikodym Theorem has a number of applications. In probability theory, the following application is of particular importance.

Example (Conditional Expectation)

Let μ be a finite measure on the σ-algebra $\mathcal{A}$ and $h \geq 0$ be a μ-integrable function. Then the measure ν, given as $d\nu = h \, d\mu$, is finite, too. Let moreover $\mathcal{A}'$ be a σ-algebra contained in $\mathcal{A}$. Restricting μ and ν to $\mathcal{A}'$ we obtain finite measures μ' and ν'. Since $\nu \ll \mu$ we have $\nu' \ll \mu'$. By the Radon-Nikodym Theorem, there exists a $\mathcal{A}'$-measurable function $h' \geq 0$ such that $d\nu' = h' \, d\mu'$. This means that

$$\int_{A'} h \, d\mu = \int_{A'} h' \, d\mu$$

holds for all $A' \in \mathcal{A}'$. We thus have adapted the measurability of the density to the σ-algebra $\mathcal{A}'$. In probability theory, h' is called the *conditional expectation of* h *given* $\mathcal{A}'$; it is μ-a.e. unique. The case of an arbitrary μ-integrable function h can be treated by decomposing it into its positive and negative part. In Chap. 12 we will make aquaintance with a different approach to conditional expectations which is based on the completeness of the space $L_2(\mu)$ instead of the Radon-Nikodym Theorem.

Another application of the theorem is concerned with the decomposition of a measure into an absolutely continuous and a singular part.

Proposition 9.3 (Lebesgue Decomposition). *Let μ and ν be σ-finite measures on a σ-algebra $\mathcal{A}$. Then there exist measures μ_a and μ_s with the properties:*

(i) $\mu = \mu_a + \mu_s$,
(ii) $\mu_a \ll \nu$ and $\mu_s \perp \nu$.

μ_a and μ_s are uniquely determined.

Proof. Obviously ν is absolutely continuous w.r.t. the measure $\mu + \nu$. According to the Radon-Nikodym Theorem, ν has a density $h \geq 0$ w.r.t. $\mu + \nu$, that is,

$$\nu(A) = \int_A h \, d\mu + \int_A h \, d\nu$$

holds for any $A \in \mathcal{A}$. We set

$$\mu_a(A) := \mu\big(A \cap \{h > 0\}\big) , \quad \mu_s(A) := \mu\big(A \cap \{h = 0\}\big) .$$

Then (i) obviously holds. If A is a ν-null set, $\int_A h \, d\mu = 0$ follows. Therefore, we have $h1_A = 0$ μ-a.e. resp. $1_{A\cap\{h>0\}} = 0$ μ-a.e. or $\mu\big(A \cap \{h > 0\}\big) = 0$. This shows that $\mu_a \ll \nu$. In addition, $\mu_s(h > 0) = 0$ and $\nu(h = 0) = \int_{\{h=0\}} h \, d(\mu + \nu) = 0$, therefore $\mu_s \perp \nu$ holds.

Let now $\mu = \mu'_a + \mu'_s$ be another decomposition with the properties (i) and (ii). Then there are measurable sets N, N' satisfying $\mu_s(N) = \mu'_s(N') = 0$, and whose complements are ν-null sets. Therefore we get $\mu_a(N^c) = \mu_a\big((N')^c\big) = 0$. For measurable sets A it follows that

$$\mu_a(A) = \mu_a(A \cap N \cap N') = \mu(A \cap N \cap N') .$$

An analogous equality holds for μ'_a, and so $\mu_a = \mu'_a$.

In the case $\mu(A) < \infty$ we obtain from (i) that $\mu_s(A) = \mu'_s(A)$. Since μ is assumed to be σ-finite, we also get $\mu_s = \mu'_s$. □

A Singular Measure on the Cantor Set*

We consider measures μ which are singular with respect to the Lebesgue measure λ on $\mathbb{R}$. An example is given by the Dirac measure $\mu = \delta_x$ whose whole mass is concentrated in $x \in \mathbb{R}$. A point x for which $\mu(\{x\}) > 0$ is called an *atom* of μ. Discrete measures composed from countably many atoms are obviously singular with respect to the Lebesgue measure. It is less obvious that there also exist measures which are singular to the Lebesgue measure but do not possess atoms.

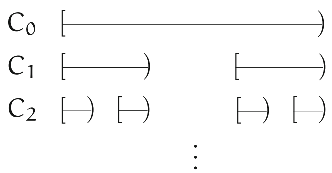
In order to construct such a measure we consider a variant of the Cantor set,[4] a subset C of the semi-open interval $[0, 1)$ within $\mathbb{R}$. Geometrically C is easily accessible: one decomposes the interval $C_0 := [0, 1)$ into parts $[0, 1/3)$, $[1/3, 2/3)$ and $[2/3, 1)$ of equal length and removes the middle part:

$$C_1 := [0, 1/3) \cup [2/3, 1) \, .$$

With the two remaining intervals one proceeds analogously:

$$\begin{aligned} C_2 &:= [0, 1/9) \cup [2/9, 1/3) \cup [2/3, 7/9) \cup [8/9, 1) \\ &= \bigcup_{a_1 \in \{0,2\}} \bigcup_{a_2 \in \{0,2\}} \big[a_1/3 + a_2/9, a_1/3 + a_2/9 + 1/9\big) \, . \end{aligned}$$

In a picture this looks as



Having removed the middle interval n times we arrive at the set

$$C_n := \bigcup_{a_1 \in \{0,2\}} \cdots \bigcup_{a_n \in \{0,2\}} \Big[\sum_{k=1}^{n} a_k 3^{-k}, \sum_{k=1}^{n} a_k 3^{-k} + 3^{-n}\Big) \, ,$$

thus $C_1 \supset C_2 \supset \cdots$. We define our version of the *Cantor set* as the result after ∞ many steps,

$$C := \bigcap_{n=1}^{\infty} C_n \, .$$

(If we construct C from closed instead of semi-open intervals, as is the common procedure, we obtain the usual Cantor set which moreover is compact. Here such subtleties are irrelevant; our procedure avoids having to deal with nonunique b-nary representations of numbers.)

[4]Georg Cantor, 1845–1918, born in St. Petersburg, active in Halle. He was the founder of set theory.

C is a null set, indeed we always remove one third, so that $\lambda(C_{n+1}) = \frac{2}{3}\lambda(C_n)$. It follows that $\lambda(C_n) = (2/3)^n$ and

$$\lambda(C) = 0 \, .$$

To describe C more precisely, we utilize the b-nary representation (to the basis $b = 2, 3, \ldots$)

$$x = \sum_{k=1}^{\infty} x_k b^{-k}$$

of numbers $x \in [0, 1)$. We assume that the sequence $x_1, x_2, \ldots$ belongs to

$$\mathcal{D}_b := \left\{(x_k)_{k\geq 1} : x_k \in \{0, 1, \ldots, b-1\} \, , \; x_k \neq b-1 \; \infty\text{-often}\right\} .$$

As we know, in this way we achieve uniqueness in the representation of x. Then $[0, 1/3)$, $[1/3, 2/3)$, $[2/3, 1)$ are the regions for which x in ternary representation ($b = 3$) has the coefficient x_1 equal to 0, to 1, and to 2. Thus it holds that

$$C_1 = \left\{ \sum_{k\geq 1} x_k 3^{-k} : (x_k)_{k\geq 1} \in \mathcal{D}_3 \, , \; x_1 \neq 1 \right\}$$

and iteratively

$$C_n = \left\{ \sum_{k\geq 1} x_k 3^{-k} : (x_k)_{k\geq 1} \in \mathcal{D}_3 \, , \; x_1, \ldots, x_n \neq 1 \right\}$$

and finally

$$C = \left\{ \sum_{k\geq 1} x_k 3^{-k} : (x_k)_{k\geq 1} \in \mathcal{D}_3 \, , \; x_1, x_2, \ldots \neq 1 \right\} .$$

Therefore C not only is nonempty, but has the same cardinality as the interval $[0, 1)$: the assignment

$$y := \sum_{k=1}^{\infty} y_k 2^{-k} \;\leftrightarrow\; \sum_{k=1}^{\infty} 2y_k 3^{-k} =: \varphi(y) \, , \quad (y_k)_{k\geq 1} \in \mathcal{D}_2 \, ,$$

gives rise to a bijection $\varphi : [0, 1) \to C$. It is strictly monotone, since $y < y'$ holds if and only if there exists an n such that $y_n < y'_n$ and $y_k = y'_k$ for any $k < n$, and this implies that $\varphi(y) < \varphi(y')$.

The singular measure μ we are looking for can now be found as the image measure of the Lebesgue measure (restricted to $[0, 1)$) under the mapping φ, so

$$\mu(B) := \lambda\big(\varphi^{-1}(B)\big)$$

for Borel sets $B \subset [0, 1)$. As λ does not possess atoms and φ is injective, μ too has no atoms. Their mutual singularity follows from $\lambda(C) = 0$, $\mu(C^c) = 0$.

Differentiability*

We now change over to consider functions $f : [a, b] \to \mathbb{R}$. We want to ascertain which functions admit an integral representation $f(x) = f(a) + \int_a^x h(z)\, dz$. It is natural to obtain h by differentiating f. Therefore, initially we concern ourselves with differentiation namely of monotone functions.

Proposition 9.4 (Lebesgue). *Let* $a < b$ *be real numbers and let* $f : [a, b] \to \mathbb{R}$ *be an increasing function. Then* f *is differentiable a.e.* (*w.r.t. the Lebesgue measure*), *and there exists a measurable function* $f' : [a, b] \to \mathbb{R}_+$ *such that* $f'(x)$, *for almost every* $x \in (a, b)$, *is equal to the derivative of* f *at the point* x. *Moreover it holds that*

$$\int_a^b f'(z)\, dz \le f(b) - f(a) .$$

The proof rests on comparing, for any $a < x < b$, the following four “right and left sided, upper and lower” derivatives:

$$f'_{ro}(x) := \limsup_{h\downarrow 0} \frac{f(x+h) - f(x)}{h} , \quad f'_{ru}(x) := \liminf_{h\downarrow 0} \frac{f(x+h) - f(x)}{h} ,$$

$$f'_{lo}(x) := \limsup_{h\downarrow 0} \frac{f(x) - f(x-h)}{h} , \quad f'_{lu}(x) := \liminf_{h\downarrow 0} \frac{f(x) - f(x-h)}{h} .$$

Since f is monotone, all four expressions are nonnegative. Differentiability in x means that they have the same finite value.

The question of measurability does not create problems: Since f is monotone, we have $\sup_{h\in(0,r]}(f(x+h) - x)/h = \sup_{h\in(0,r]\cap\mathbb{Q}}(f(x+h) - x)/h$, and so

$$f'_{ro}(x) = \lim_{n\to\infty} \sup_{h\in(0,n^{-1}]\cap\mathbb{Q}} \frac{f(x+h) - f(x)}{h} .$$

From the usual properties of measurable functions we obtain the Borel measurability of $f'_{ro} : (a,b) \to \bar{\mathbb{R}}_+$ and in the same manner that of f'_{ru}, f'_{lo} and f'_{lu}. The Borel measurability of the set D_f of all points $x \in (a,b)$ where f is differentiable follows from the fact that

$$D_f = \left\{x \in (a,b) : f'_{lu}(x) = f'_{lo}(x) = f'_{ru}(x) = f'_{ro}(x) < \infty\right\} .$$

The proof of the remaining part of Lebesgue's Theorem is more difficult. By means of a simple particular case we want to make it plausible that discrepancies between the derivatives lead to a contradiction if they are too large. Let us assume that there are numbers $r < s$ such that $f'_{ru}(x) < r < s < f'_{lo}(x)$ for all $x \in (a,b)$. Thus for each x there exists an $h > 0$ satisfying $f(x+h) - f(x) \le rh$. Therefore it seems natural that we can find a partition $a = x_0 < x_1 < \cdots < x_{m-1} < x_m = b$ with $f(x_j) - f(x_{j-1}) \le r(x_j - x_{j-1})$ for every $j = 1, \ldots, m$. We then would have decomposed $[a,b]$ into intervals $I_j = (x_{j-1}, x_j)$ where f has only small increments, and could conclude that

$$f(b) - f(a) \le r(b-a) .$$

But using the other part of the assumption in the same manner we could find another partition $a = y_0 < y_1 < \cdots < y_{n-1} < y_n = b$ with $f(y_j) - f(y_{j-1}) \ge s(y_j - y_{j-1})$ for every $j = 1, \ldots, n$, a decomposition into intervals I'_j of larger increments of f, and we would obtain

$$f(b) - f(a) \ge s(b-a) .$$

All in all this yields a contradiction.

These considerations may be transferred in a similar manner to subintervals and to the other derivatives. This makes it plausible that contradictions can be avoided only if $f'_{lu}, f'_{ru}, f'_{lo}$ and f'_{ro} coincide almost everywhere. In what follows we want to elaborate this argument, however in general the choice of suitable intervals of smaller or bigger increments of f is a bit more complicated. We prepare this step by the following lemma concerning *Vitali coverings* of Borel sets.

Lemma (Vitali's Covering Lemma). *Let* $B \subset (a,b)$ *be a Borel set, and let* $\mathcal{V}$ *be a set of intervals* $I \subset (a,b)$ *with* $\lambda(I) > 0$ *having the property: For each* $x \in B$ *and each* $\varepsilon > 0$ *there exists an* $I \in \mathcal{V}$ *such that* $x \in I$ *and* $\lambda(I) \le \varepsilon$. *Then for every* $\varepsilon > 0$ *there exist finitely many disjoint intervals* $I_1, \ldots, I_n \in \mathcal{V}$ *such that*

$$\lambda\Big(B \setminus \bigcup_{j=1}^{n} I_j\Big) \le \varepsilon .$$

Proof. We inductively construct the intervals $I_1, I_2, \ldots \in \mathcal{V}$. I_1 is chosen arbitrarily from $\mathcal{V}$. Having chosen $I_1, \ldots, I_k$, we set

$$s_k := \sup \left\{ \lambda(I) \,:\, I \in \mathcal{V}, I \subset (a,b) \setminus \bigcup_{j=1}^{k} I_j \right\} .$$

If $B \subset \bigcup_{j=1}^{k} \bar{I}_j$ ($\bar{I}_j$ being the topological closure of I_j), the construction terminates, otherwise $s_k > 0$ holds by the assumption of the lemma. We then choose $I_{k+1} \in \mathcal{V}$ such that $\lambda(I_{k+1}) \geq s_k/2$.

If the construction terminates after n steps, the intervals $I_1, \ldots, I_n$ obviously satisfy our assertion. If the construction does not terminate, since the intervals are disjoint we get

$$\sum_{j=1}^{\infty} \lambda(I_j) = \lambda\Big(\bigcup_{j=1}^{\infty} I_j\Big) \leq b - a < \infty .$$

It follows that $\lambda(I_k) \to 0$ and that $s_k \to 0$ when $k \to \infty$. In addition, for any $\varepsilon > 0$ there exists a natural number n such that $\sum_{l>n} \lambda(I_l) \leq \varepsilon/5$. We show that for this n the assertion of the lemma holds.

To this end we prove that

$$B \setminus \bigcup_{j=1}^{n} \bar{I}_j \subset \bigcup_{l>n} I_l^* ,$$

where I_l^* denotes the interval which has the same midpoint as I_l, but is 5 times longer. Let $x \in B \setminus \bigcup_{j=1}^{n} \bar{I}_j$ be arbitrary. Since $\bigcup_{j=1}^{n} \bar{I}_j$ is closed, there exists an $I \in \mathcal{V}$ with $x \in I$ such that $I, I_1, \ldots, I_n$ are disjoint intervals. If I would be disjoint with all the intervals I_k, it would follow that $\lambda(I) \leq s_k$ for all k and therefore $\lambda(I) = 0$, a contradiction. Thus there exists an $l > n$ such that $I \cap I_l \neq \varnothing$ and $I \cap I_j = \varnothing$ for all $j < l$. It follows that $\lambda(I) \leq s_{l-1} \leq 2\lambda(I_l)$. From this and since $I \cap I_l \neq \varnothing$ we get that I_l, stretched by a suitable factor, covers the interval I. More precisely, $I \subset I_l^*$ holds for the interval I_l^* of 5-fold length defined above. As $x \in I$, $x \in I_l^*$ follows. This yields the assertion.

We conclude that

$$\lambda\Big(B \setminus \bigcup_{j=1}^{n} I_j\Big) \leq \sum_{l>n} \lambda(I_l^*) = 5 \sum_{l>n} \lambda(I_l) \leq \varepsilon .$$

The lemma is proved. □

Proof of Lebesgue's Theorem. Let $r < s$ be real numbers. The main part of the proof consists in showing that

$$N_{rs} := \{x \in (a, b) : f'_{ru}(x) < r < s < f'_{lo}(x)\}$$

is a Lebesgue null set.

Let $\varepsilon > 0$. Due to the outer regularity of the Lebesgue measure according to Proposition 7.5, there exists an open set O such that $N_{rs} \subset O \subset (a, b)$ and $\lambda(O) \leq \lambda(N_{rs}) + \varepsilon$. We consider the system $\mathcal{V}$ of all intervals $(x, x + h) \subset O$ satisfying $x \in N_{rs}$, $h > 0$ and $f(x + h) - f(x) \leq rh$. By the definition of N_{rs}, the system $\mathcal{V}$ fulfils the conditions of Vitali's Covering Lemma for $B = N_{rs}$, therefore there exist disjoint intervals $I_1 = (x_1, x_1 + h_1), \ldots, I_m = (x_m, x_m + h_m)$ such that

$$\lambda\Big(N_{rs} \setminus \bigcup_{j=1}^{m} I_j\Big) \leq \varepsilon$$

and

$$\sum_{j=1}^{m} \big(f(x_j + h_j) - f(x_j)\big) \leq r \sum_{j=1}^{m} h_j = r\lambda\Big(\bigcup_{j=1}^{m} I_j\Big) \leq r\lambda(O) \leq r\big(\lambda(N_{rs}) + \varepsilon\big) .$$

We moreover consider the system $\mathcal{V}'$ of all intervals $(y - k, y) \subset \bigcup_{j=1}^{m} I_j$ with $y \in N_{rs}$, $k > 0$ and $f(y) - f(y - k) \geq sk$. The system $\mathcal{V}'$, too, by the definition of N_{rs} satisfies the conditions of the lemma for $B = N_{rs} \cap \bigcup_{j=1}^{m} I_j$, thus there exist disjoint intervals $I'_1 = (y_1 - k_1, y_1), \ldots, I'_n = (y_n - k_n, y_n)$ such that

$$\lambda\Big(\big(N_{rs} \cap \bigcup_{j=1}^{m} I_j\big) \setminus \bigcup_{l=1}^{n} I'_l\Big) \leq \varepsilon$$

and

$$\sum_{j=1}^{n} \big(f(y_j) - f(y_j - k_j)\big) \geq s \sum_{j=1}^{n} k_j = s\lambda\Big(\bigcup_{l=1}^{n} I'_l\Big) \geq s\big(\lambda(N_{rs}) - 2\varepsilon\big) .$$

Since each I'_l is contained in one of the I_j, and since f is monotone,

$$\sum_{j=1}^{n} \big(f(y_j) - f(y_j - k_j)\big) \leq \sum_{j=1}^{m} \big(f(x_j + h_j) - f(x_j)\big) .$$

Altogether we get $s\big(\lambda(N_{rs}) - 2\varepsilon\big) \leq r\big(\lambda(N_{rs}) + \varepsilon\big)$. Since $r < s$ we obtain, letting $\varepsilon \to 0$, that $\lambda(N_{rs}) = 0$, as claimed.

As the rational numbers are dense in $\mathbb{R}$, we have

$$\{x \in (a,b) : f'_{ru}(x) < f'_{lo}(x)\} = \bigcup_{r,s\in\mathbb{Q},r<s} N_{rs} ,$$

and by virtue of σ-subadditivity we get $\lambda(f'_{ru} < f'_{lo}) = 0$, that is $f'_{lo} \leq f'_{ru}$ a.e. In the same manner $f'_{ro} \leq f'_{lu}$ a.e. follows (interchanging intervals to the right resp. left). In addition, $f'_{ru} \leq f'_{ro}$ and $f'_{lu} \leq f'_{lo}$ obviously hold, and we obtain

$$f'_{lo} \leq f'_{ru} \leq f'_{ro} \leq f'_{lu} \leq f'_{lo} \quad \text{a.e.}$$

Thus the four derivatives are equal a.e., and f is a.e. differentiable; it may happen that the derivatives have the value ∞.

In order to show that the derivatives are finite a.e., we consider

$$f_n(x) := n\big(f(x+1/n) - f(x)\big)1_{(a,b-1/n)}(x) .$$

We have $\lim_{n\to\infty} f_n(x) = f'_{ro}(x)$ a.e. By Fatou's Lemma and monotonicity, it follows that

$$\begin{aligned}\int_a^b f'_{ro}(z)\,dz &\leq \liminf_{n\to\infty} \int_a^b f_n(z)\,dz \\ &= \liminf_{n\to\infty}\Big(n\int_{b-1/n}^{b} f(z)\,dz - n\int_a^{a+1/n} f(z)\,dz\Big) \leq f(b) - f(a)\end{aligned}$$

Consequently, $f'_{ro} < \infty$ a.e., and f possesses an a.e. finite derivative. Setting $f'(x) := f'_{ro}(x)$ when $x \in D_f$ and $f'(x) := 0$ otherwise, the assertion follows. □

Example (Cantor's Function)

In the previous section we have constructed a strictly increasing function φ from $[0,1)$ onto the Cantor set C. Its inverse function $\psi : C \to [0,1)$ can be extended to a monotone function $f : [0,1) \to [0,1)$. For this purpose we recall that $[0,1) \setminus C$ consists of countably many disjoint intervals $[a_n, b_n)$. For $x \in [a_n, b_n)$ we set $f(x) := f(b_n)$. Then f is monotone and surjective, and this implies the continuity of f.

Obviously, $f'(x) = 0$ for all $x \in (a_n, b_n)$. As C is a null set, it follows that $f' = 0$ a.e. Hence, in this example we have that $\int_0^1 f'(z)\,dz < f(1) - f(0)$.

There are even *strictly* monotone, continuous functions $f : [0,1) \to [0,1)$ whose derivative vanishes a.e. Such functions are more difficult to construct.

Absolutely Continuous Functions*

Now we want to characterize those monotone functions for which in the foregoing proposition concerning derivatives of monotone functions we even have equality

$f(x) = f(a) + \int_a^x f'(z)\,dz$. To this end we introduce the following notion (not restricted to monotone functions) which strengthens the notions of continuity and uniform continuity.

Definition

A function $f : [a, b] \to \mathbb{R}$ is called *absolutely continuous*, if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $a \le x_1 < y_1 \le x_2 < y_2 \le \cdots \le x_n < y_n \le b$ it holds that

$$\sum_{i=1}^{n}(y_i - x_i) \le \delta \quad \Rightarrow \quad \sum_{i=1}^{n}\big|f(y_i) - f(x_i)\big| \le \varepsilon\,.$$

Lipschitz continuous functions, for example, are absolutely continuous. These are functions f for which there exists an $L < \infty$ such that $|f(x)-f(y)| \le |x-y|$ holds for all x, y. Functions whose derivative exists everywhere and is bounded belong to this class.

Proposition 9.5. *A monotone increasing function* $f : [a, b] \to \mathbb{R}$ *is absolutely continuous if and only if there exists a nonnegative Lebesgue integrable function* $h : [a, b] \to \mathbb{R}$ *such that*

$$f(x) = f(a) + \int_a^x h(z)\,dz\,.$$

In this case $h(x) = f'(x)$ *holds for almost all* $x \in (a, b)$.

Proof. (i) First, let us assume that f possesses the stated integral representation. Then for any $a \le x_1 < y_1 \le x_2 < y_2 \le \cdots \le x_n < y_n \le b$ and $c > 0$ it holds, setting $A := \bigcup_{i=1}^{n}[x_i, y_i]$, that

$$\sum_{i=1}^{n}\big|f(y_i) - f(x_i)\big| = \int_A h(z)\,dz \le c\lambda(A) + \int_{\{h>c\}} h(z)\,dz\,.$$

For any given $\varepsilon > 0$ we choose c large enough such that the rightmost integral becomes smaller than $\varepsilon/2$. If now $\sum_{i=1}^{n}(y_i - x_i) = \lambda(A) \le \delta$ with $\delta := \varepsilon/(2c)$, we get that $\sum_{i=1}^{n}\big|f(y_i) - f(x_i)\big| \le \varepsilon$. Therefore, f is absolutely continuous.

(ii) Next we show that f a.e. has the derivative h if we assume the integral representation to hold. By the results of the previous section, f is differentiable a.e., thus

$$f_n(x) := n\big(f(x + 1/n) - f(x)\big) \cdot 1_{(a,b-1/n)}(x)$$

converges a.e. to $f'(x) \ge 0$. We have to prove that $f' = h$ a.e.

We first consider the case where $h(x) \le c$ for some $c < \infty$ and all x. Then $0 \le f_n(x) \le c$ follows, and the dominated convergence theorem yields for $a < x < b$ that

$$\begin{aligned}\int_a^x f'(z)\,dz &= \lim_{n\to\infty} \int_a^x n\big(f(z+1/n) - f(z)\big)\,dz \\ &= \lim_{n\to\infty} \left(n \int_x^{x+1/n} f(z)\,dz - n \int_a^{a+1/n} f(z)\,dz\right) \\ &= f(x) - f(a) = \int_a^x h(z)\,dz\,.\end{aligned}$$

This means that the two measures on $[a, b]$, given by the densities $f'\,d\lambda$ and $h\,d\lambda$, coincide on all subintervals of $[a, b]$. These intervals form a $\cap$-stable generator of the Borel σ-algebra, therefore the two measures, too, coincide according to the uniqueness theorem. Thus the densities f' and h are equal a.e.

The general case now can be treated using the decomposition

$$f(x) - f(a) = f_1(x) + f_2(x) := \int_a^x h_1(z)\,dz + \int_a^x h_2(z)\,dz\,,$$

where $h_1 := h1_{\{h\le c\}}$, $h_2 := h1_{\{h>c\}}$, and $c > 0$ is given. f_2 is monotone increasing and thus has a.e. a nonnegative derivative. As h_1 is bounded by c, from what we just proved it follows that $h_1(x) = f_1'(x) \le f'(x)$ a.e. Since h is finite a.e., letting $c \to \infty$ we obtain that $h \le f'$ a.e. On the other hand, the proposition concerning the differentiation of monotone functions implies that

$$\int_a^b h(z)\,dz = f(b) - f(a) \ge \int_a^b f'(z)\,dz\,.$$

All in all this yields $h = f'$ a.e. and thus the assertion.

(iii) Finally, let f be absolutely continuous. We have to prove that f has the integral representation as stated. To this end we will show that the function

$$g(x) := f(x) - \int_a^x f'(z)\,dz$$

has the constant value $f(a)$.

By what we have seen so far g has the following properties: by virtue of the proposition concerning the derivatives of monotone functions we have $\int_x^y f'(z)\,dz \le f(y) - f(x)$ for $x < y$, therefore g is monotonically increasing. It then follows that $|g(x) - g(y)| \le |f(x) - f(y)|$, thus g is absolutely continuous because so is f. Finally, by (ii) we have $g'(x) = f'(x) - f'(x)$ a.e., that is, the derivative of g vanishes a.e.

Let B be the Borel set of all $x \in (a, b)$ with $g'(x) = 0$, and let $\varepsilon > 0$. We consider the system $\mathcal{V}$ of all intervals $[y, z] \subset (a, b)$ satisfying $y < z$ and $g(z) - g(y) \leq \varepsilon(z - y)$. For each $x \in B$ and each $\delta > 0$ there exists an interval $I \in \mathcal{V}$ with $x \in I$ and $\lambda(I) \leq \delta$. By Vitali's Covering Lemma, for every $\delta > 0$ we can find disjoint intervals $I_j = [y_j, z_j] \in \mathcal{V}$ such that $\lambda\big(B \setminus \bigcup_{j=1}^{n} I_j\big) \leq \delta$. Since $\lambda\big([a, b] \setminus B\big) = 0$, this means that

$$(y_1 - a) + \sum_{i=1}^{n-1} (y_{i+1} - z_i) + (b - z_n) \leq \delta .$$

If we choose δ (depending on ε) sufficiently small, the absolute continuity of g implies that

$$g(y_1) - g(a) + \sum_{i=1}^{n-1} \big(g(y_{i+1}) - g(z_i)\big) + g(b) - g(z_n) \leq \varepsilon .$$

According to the definition of the intervals I_j we moreover have that

$$\sum_{j=1}^{n} \big(g(z_j) - g(y_j)\big) \leq \sum_{j=1}^{n} \varepsilon(z_j - y_j) \leq \varepsilon(b - a) .$$

Adding the previous two inequalities yields that $g(b) - g(a) \leq \varepsilon + \varepsilon(b - a)$, and letting $\varepsilon \to 0$ we obtain $g(b) \leq g(a) = f(a)$. Since on the other hand g increases monotonically, it follows that $g(x) = g(a) = f(a)$ for all $x \in [a, b]$. This is the desired integral representation. □

Functions of Bounded Variation*

We now want to drop the assumption of monotonicity, which played an important role in the previous two sections, and pass to functions which can be represented as differences of monotone functions.

Definition

A function $f : [a, b] \to \mathbb{R}$ is said to have *bounded variation* (or *finite variation*), if there exists a $c > 0$ such that for all $n \in \mathbb{N}$ and all partitions $a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b$ of length n,

$$\sum_{i=1}^{n} \big|f(x_i) - f(x_{i-1})\big| \leq c .$$

Proposition 9.6 (Jordan Decomposition). *A function* $f : [a, b] \to \mathbb{R}$ *has bounded variation if and only if it is equal to the difference of two monotonically increasing functions* $f_1, f_2 : [a, b] \to \mathbb{R}$*:*

$$f = f_1 - f_2 .$$

Proof. First, let f be the difference of the monotonically increasing functions f_1, f_2. Then

$$\sum_{i=1}^{n} \left|f(x_i) - f(x_{i-1})\right| \leq \sum_{i=1}^{n} \big(f_1(x_i) - f_1(x_{i-1})\big) + \sum_{i=1}^{n} \big(f_2(x_i) - f_2(x_{i-1})\big)$$
$$= f_1(b) - f_1(a) + f_2(b) - f_2(a) .$$

Thus f has bounded variation.

Conversely, assume that f has bounded variation. For $a \leq y < z \leq b$ the nonnegative quantity

$$v(y, z) := \sup_{y=x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = z} \sum_{i=1}^{n} \left|f(x_i) - f(x_{i-1})\right|$$

is termed the *variation* of f on the interval $[y, z]$. Obviously, it is finite for functions of bounded variation. If $y < u < z$, we may always adjoin u to the partition $x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n$, because the corresponding sums become larger and the supremum remains unchanged. Since we may select the partition below and above u separately, it follows that

$$v(y, z) = v(y, u) + v(u, z) .$$

We set

$$f_1(y) := v(a, y) , \quad f_2(y) := v(a, y) - f(y) ,$$

thus $f_1 - f_2 = f$. For $y < z$ it holds that $f_1(z) - f_1(y) = v(y, z) \geq 0$ and

$$f_2(z) - f_2(y) = v(y, z) - f(z) + f(y) \geq v(y, z) - |f(z) - f(y)| \geq 0 .$$

Therefore, f_1 and f_2 are monotonically increasing. □

According to the proposition on differentiation of monotone functions, every function of bounded variation can be differentiated a.e. For absolute continuous functions the following stronger result is valid.

Proposition 9.7. *Any absolutely continuous function* $f : [a, b] \to \mathbb{R}$ *can be represented as the difference* $f = f_1 - f_2$ *of two monotonically increasing, absolutely continuous functions* f_1, f_2.

Proof. As in the previous proof, we work with the variation $v(y, z)$. Absolute continuity of f means that for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $v(y, z) \leq \varepsilon$ for $z - y \leq \delta$. Since $v(y, z) = v(y, u) + v(u, z)$, it follows that $v(y, z) \leq n\varepsilon$ for $z - y \leq n\delta$ and all $n \in \mathbb{N}$. In particular, $v(y, z) < \infty$ holds for all $a \leq y < z \leq b$. Absolutely continuous functions therefore have bounded variation.

We proceed as in the previous proof and obtain monotone functions $f_1(y) := v(a, y)$, $f_2(y) := v(a, y) - f(y)$, such that $f = f_1 - f_2$. It remains to show that f_1 (and therefore $f_2 = f_1 - f$) is absolutely continuous. Let $\delta, \varepsilon > 0$ and $a \leq y_1 < z_1 \leq y_2 < z_2 \leq \cdots \leq y_n < z_n \leq b$ such that $\sum_{j=1}^{n}(z_i - y_i) \leq \delta$. By the definition of the supremum, there exist partitions $y_i = x_{i,0} \leq x_{i,1} \leq \cdots \leq x_{i,n_i} = z_i$ such that

$$v(y_i, z_i) \leq 2 \sum_{j=1}^{n_i} \left| f(x_{i,j}) - f(x_{i,j-1}) \right| .$$

We get

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} (x_{i,j} - x_{i,j-1}) = \sum_{i=1}^{n} (z_i - y_i) \leq \delta .$$

Due to the absolute continuity of f it follows that

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left| f(x_{i,j}) - f(x_{i,j-1}) \right| \leq \frac{\varepsilon}{2} ,$$

if δ is sufficiently small. We obtain

$$\sum_{i=1}^{n} \big(f_1(z_i) - f_1(y_i) \big) = \sum_{i=1}^{n} v(y_i, z_i) \leq \varepsilon ,$$

thus f_1 is absolutely continuous as claimed. □

Generalizing from the situation of monotone functions, we present the following characterization of absolutely continuous functions.

Proposition 9.8. *A function* $f : [a, b] \to \mathbb{R}$ *is absolutely continuous if and only if there exists a Lebesgue integrable function* $h : [a, b] \to \mathbb{R}$ *such that*

$$f(x) = f(a) + \int_a^x h(z)\,dz\,.$$

In this case $h(x) = f'(x)$ *for almost all* $x \in (a, b)$.

Proof. If f is absolutely continuous, we have $f = f_1 - f_2$ for some monotone absolutely continuous function f_1, f_2. For those functions $f_i(x) = f_i(a) + \int_a^x h_i(z)\,dz$, and we obtain the integral representation for f setting $h := h_1 - h_2$.

If, conversely, the integral representation holds, it follows that $f = f_1 - f_2$ with the monotone functions $f_1(x) := f(a) + \int_a^x h^+(z)\,dz$, $f_2(x) := \int_a^x h^-(z)\,dz$. The functions f_1 and f_2 are absolutely continuous, and thus so is f.

The final claim results from the corresponding assertions for f_1 and f_2. □

Signed Measures*

When treating measures, one may also drop the monotonicity requirement, similarly as we did for functions in the previous section.

Definition

A mapping $\delta : \mathcal{A} \to \bar{\mathbb{R}}$ from a σ-algebra on a measurable space $(S, \mathcal{A})$ to $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ is called a *signed measure*, if $\delta(\emptyset) = 0$ and if for every (finite or infinite) sequence of disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$ one has that

$$\delta\Big(\bigcup_{n\ge1} A_n\Big) = \sum_{n\ge1} \delta(A_n)\,.$$

It is part of the definition that the sum on the right-hand side is always well-defined. On one hand, this means that the sum does not depend on the order of summation. On the other hand, ∞ and $-\infty$ are not allowed to appear both during summation. This excludes the possibility that there are two sets $A, A' \in \mathcal{A}$ with $\delta(A) = \infty$ and $\delta(A') = -\infty$. (Then $\delta(A \cap A')$ would have to be finite, and the disjoint sets $A \setminus A'$ and $A' \setminus A$ would have the value ∞ and $-\infty$.) Therefore either ∞ or $-\infty$ is not present among the values of δ.

Obviously, a signed measure arises when one considers the difference $\delta = \mu - \nu$ of two measures, at least one of them being finite. It turns out that one obtains all signed measures in this manner. More precisely, the following proposition holds.

Proposition 9.9 (Jordan Decomposition of Signed Measures). *Let δ be a signed measure. Then there are measures δ^+ und δ^-, at least one of them being finite, such that $\delta = \delta^+ - \delta^-$ and $\delta^+ \perp \delta^-$. These measures are uniquely determined, and it holds that*

$$\delta^+(A) = \sup_{A' \subset A} \delta(A') , \quad \delta^-(A) = -\inf_{A' \subset A} \delta(A') .$$

δ^+ and δ^- are termed *positive* and *negative variation* of δ. Thus one may think of a signed measure as a charge distribution in the space S, with positive and negative charges (as one may think of a measure as a mass distribution in the space). The proof of the proposition is based on the Hahn decomposition for signed measures.

Proposition 9.10 (Hahn Decomposition). *Let δ be a signed measure on a σ-algebra. Then there are measurable sets $A_\geq$ and $A_\leq = S \setminus A_\geq$ such that for all measurable sets A we have*

$$\delta(A) \geq 0 \text{ for } A \subset A_\geq ,$$
$$\delta(A) \leq 0 \text{ for } A \subset A_\leq .$$

Proof. It suffices to treat the case where $\delta(A) > -\infty$ for all measurable A. In that case we may carry over completely the proof of Proposition 9.2. □

Proof of the Jordan decomposition. Letting $A_\geq, A_\leq$ be a Hahn decomposition of δ, we set

$$\delta^+(A) := \delta(A \cap A_\geq) , \quad \delta^-(A) := -\delta(A \cap A_\leq) .$$

δ^+ and δ^- satisfy $\delta = \delta^+ - \delta^-$ and $\delta^+ \perp \delta^-$.

Concerning uniqueness: Let $\delta = \mu - \nu$ and $\mu \perp \nu$. For measurable sets $A' \subset A$ then

$$\delta(A') \leq \mu(A') \leq \mu(A)$$

holds. Moreover, there exists a measurable set B such that $\nu(B) = \mu(B^c) = 0$. It follows that

$$\delta(A \cap B) = \mu(A \cap B) = \mu(A) .$$

Taken together the two assertions yield

$$\mu(A) = \sup_{A' \subset A} \delta(A') .$$

Analogously,

$$\nu(A) = -\inf_{A' \subset A} \delta(A') .$$

Therefore, μ and ν are uniquely determined by δ, and these formulas are valid for δ^+ resp. δ^-, too. □

Exercises

9.1 Let μ and ν be σ-finite. Prove that $\nu \ll \mu$ is equivalent to the condition

$$\forall \varepsilon > 0 \, \exists \delta > 0 \; : \; \mu(A) \leq \delta \; \Rightarrow \; \nu(A) \leq \varepsilon .$$

Hint: The Radon-Nikodym theorem helps. From $d\nu = h\, d\mu$ it follows for all $c > 0$ that

$$\nu(A) \leq \int_{A \cap \{h \leq c\}} c\, d\mu + \int_{A \cap \{h > c\}} h\, d\mu \leq c\mu(A) + \nu(h > c) .$$

9.2 Let S be uncountable, let $\mathcal{A}$ be the σ-algebra of all $A \subset S$ which are countable or whose complement is countable, and let $h : S \to \mathbb{R}$ be a nonnegative function. We consider the measures μ and ν on $\mathcal{A}$, given by $\mu(A) := \#A$ and

$$\nu(A) := \begin{cases} \sum_{x \in A} h(x) , & \text{if A countable,} \\ \infty , & \text{otherwise.} \end{cases}$$

(i) When does $\nu \ll \mu$ hold? (ii) When has ν a density w.r.t. μ? (Compare Exercise 2.1)

9.3 Let $B \subset \mathbb{R}$ be a Borel set. Prove that for almost all $x \in B$

$$\lim_{h \downarrow 0} \frac{\lambda\big([x-h, x+h] \cap B\big)}{2h} = 1 .$$

One says that almost all elements of B are *points of density* of B.

9.4 Does the continuous function $f(x) := x \sin(1/x)$, $f(0) := 0$ have bounded variation on the interval $[0, 1]$? What about $g(x) := xf(x)$?

9.5 Let $\delta = \mu - \nu$, where μ and ν are measures (one of them being finite). Then $\delta^+(A) \leq \mu(A)$ and $\delta^-(A) \leq \nu(A)$ for all measurable A.

9.6 For a signed measure δ one defines its *variation* as the measure $|\delta| := \delta^+ + \delta^-$. Prove that

$$|\delta|(A) = \sup \Big\{ \sum_{k=1}^{n} |\delta(A_k)| : A_1, \ldots, A_n \text{ are disjoint} , \bigcup_{k=1}^{n} A_k \subset A \Big\} .$$

The Jacobi Transformation Formula

10

We have determined the volume of parallelotopes in Euclidean space with the aid of determinants in Proposition 3.4. In this chapter we present a far-reaching generalization of this issue which dates back to Jacobi.[1]

Let G, H be open subsets of $\mathbb{R}^d$, and let

$$\varphi : G \to H$$

be a C^1-diffeomorphism, that is, a bijective mapping between G and H which is continuously differentiable in both directions. For any fixed $x \in G$ we thus have

$$\varphi(x + v) = \varphi(x) + \varphi'_x(v) + o\big(|v|\big) , \tag{10.1}$$

as $v \in \mathbb{R}^d$ tends to 0. Here φ'_x denotes, for any x, a linear mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$. By the inverse function theorem, φ'_x is bijective for every x, and $\varphi'_x(v)$ is jointly continuous in x and v. The inverse mapping $\psi : H \to G$ has analogous properties, and

$$\psi'_{\varphi(x)} = (\varphi'_x)^{-1} .$$

Generalizing Proposition 3.4 we now prove the following result which goes back to Jacobi.

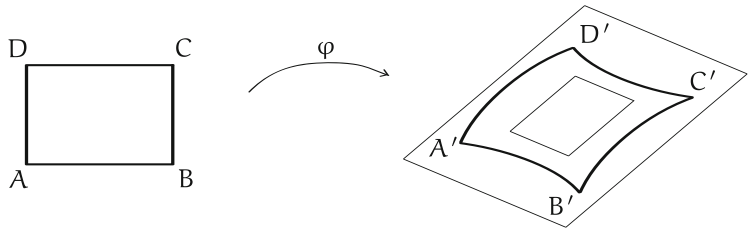
[1] Carl Gustav Jacobi, 1804–1851, born in Potsdam, active in Königsberg and Berlin. He worked in number theory, elliptic functions, and mechanics.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_10

Proposition 10.1. *For any* C^1*-diffeomorphism* $\varphi : G \to H$ *and any Borel set* $B \subset G$ *we have*

$$\lambda^d\big(\varphi(B)\big) = \int_B |\det \varphi'_x|\, dx\,.$$

Since $\varphi(B) = \psi^{-1}(B)$ and ψ is Borel measurable, $\varphi(B)$ is a Borel set. A large portion of the proof is concerned with a geometric property of diffeomorphisms, namely that the images of rectangles under φ (as shown in the following figure) can be enclosed from outside as well as from inside by parallelepipeds, in fact more and more accurately as the rectangles become smaller and smaller.



Let $Q := [-c, c)$, $c = (c_1, \ldots, c_d) \in \mathbb{R}^d_+$ be a d-dimensional interval centered at 0. Dilating it in all directions by the factor $\sigma > 0$ and translating it by $x \in \mathbb{R}^d$, we obtain the rectangle $x + \sigma Q$. Its image $\varphi(x + \sigma Q)$ can be nested using the parallelotope $\varphi'_x(\sigma Q)$. More precisely, the following fact holds.

Lemma. *Let* $K \subset G$ *be compact and* $0 < \eta < 1$. *If* $\sigma > 0$ *is sufficiently small, then*

$$\varphi(x) + (1-\eta)\varphi'_x(\sigma Q) \;\subset\; \varphi(x + \sigma Q) \;\subset\; \varphi(x) + (1+\eta)\varphi'_x(\sigma Q)$$

for all $x \in K$.

Proof. (i) As a preparation, we show that (10.1) holds uniformly on compacta. As $K \subset G$ is compact, there exists a $\kappa > 0$ such that $x + v \in G$ for any $x \in K$ and $v \in \mathbb{R}^d$ satisfying $|v| \le \kappa$. We claim that for every $\varepsilon > 0$ there exists a $\delta \in (0, \kappa]$ such that

$$\big|\varphi(x + v) - \varphi(x) - \varphi'_x(v)\big| \le \varepsilon |v| \qquad (10.2)$$

for every $x \in K$ and $v \in \mathbb{R}^d$ with $|v| \le \delta$.

To prove this we remark that the mapping $(x, w) \mapsto \varphi'_x(w)$ is continuous and hence uniformly continuous on the compact set $\{(x + v, w) : x \in K, |v| \le \kappa, |w| = 1\}$. Thus, for every $\varepsilon > 0$ there exists a $\delta \in (0, \kappa]$ such that $|\varphi'_{x+v}(w) - \varphi'_x(w)| \le \varepsilon/2$ for all $x \in K$, $|v| \le \delta$, and $|w| = 1$. For any $x \in K$, $|v| \le \delta$ we now consider the function

$$g(t) := \varphi\big(x + tv\big) - \varphi(x) - t\varphi'_x(v)\ , \quad 0 \le t \le 1\ .$$

Due to our differentiability assumptions, $dg(t)/dt = \varphi'_{x+tv}(v) - \varphi'_x(v)$. From $|g(t + h)|^2 - |g(t)|^2 = (g(t + h) - g(t)) \cdot (g(t + h) + g(t))$ (the dot denotes the scalar product) it follows that $d|g(t)|^2/dt = 2\big(\varphi'_{x+tv}(v) - \varphi'_x(v)\big) \cdot g(t)$. Using the Cauchy-Schwarz inequality we get that $\big|d|g(t)|^2/dt\big| \le \varepsilon|v||g(t)|$. Integration yields $|g(t)|^2 \le \varepsilon|v|t \sup_{0\le s\le t} |g(s)|$, thus $(\sup_{0\le t\le 1} |g(t)|)^2 \le \varepsilon|v| \sup_{0\le t\le 1} |g(t)|$ and therefore $|g(1)| \le \varepsilon|v|$. This is (10.2).

We transform (10.2) in a twofold manner. Firstly we claim that for any given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\Big|\psi'_{\varphi(x)}\big(\varphi(x + v) - \varphi(x) - \varphi'_x(v)\big)\Big| \le \varepsilon|v| \tag{10.3}$$

for all $x \in K$, $|v| \le \delta$. To show this we use that the continuous mapping $(x, w) \mapsto |\psi'_{\varphi(x)}(w)|$ attains a finite maximum m_1 on the compact set $\{(x, w) : x \in K, |w| = 1\}$. Consequently, $|\psi'_{\varphi(x)}(v)| \le m_1|v|$ for all $x \in K$, $v \in \mathbb{R}^d$, the assertion thus follows from (10.2) if we replace there ε by ε/m_1.

Secondly we claim that for every given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\Big|\psi(\varphi(x) + \varphi'_x(v)) - x - v\Big| \le \varepsilon|v| \tag{10.4}$$

for all $x \in K$, $|v| \le \delta$. To show this we use that the continuous mapping $(x, w) \mapsto |\varphi'_x(w)|$ attains a finite maximum m_2 on $\{(x, w) : x \in K, |w| = 1\}$. If $\delta > 0$ is sufficiently small, we therefore have $\varphi(x) + \varphi'_x(v) \in H$ whenever $x \in K$ and $v \in \mathbb{R}^d$ with $|v| \le \delta$. Now (10.2) yields for $\varphi(K)$, ψ, $\varphi(x)$ and $\varphi'_x(v)$ in place of K, φ, x and v, the inequality $|\psi(\varphi(x) + \varphi'_x(v)) - x - v| \le \varepsilon|\varphi'_x(v)| \le \varepsilon m_2|v|$. The claim follows if we replace ε with ε/m_2.

(ii) We now prove the assertion of the lemma.

Concerning the right inclusion: We investigate when for any $x \in K$ and any $v \in \sigma Q$ there exists a $u \in \eta\sigma Q$ such that $\varphi(x + v) = \varphi(x) + \varphi'_x(v + u)$ holds. If σ is sufficiently small, we have $x + \sigma Q \subset G$ for all $x \in K$. For any $x \in K$, $v \in \sigma Q$ we may then form

$$u := \psi'_{\varphi(x)}\big(\varphi(x + v) - \varphi(x) - \varphi'_x(v)\big)\ .$$

For any given $\varepsilon > 0$, by (10.3) one has $|u| \le \varepsilon\eta|v|$ whenever σ is sufficiently small. If ε is chosen sufficiently small, in view of the shape of rectangles and because $v \in \sigma Q$ it follows that $u \in \eta\sigma Q$, and therefore $v + u \in (1 + \eta)\sigma Q$.

According to the definition of u,

$$\varphi(x + v) = \varphi(x) + \varphi'_x(v + u) ,$$

therefore $\varphi(x + \sigma Q) \subset \varphi(x) + (1 + \eta)\varphi'_x(\sigma Q)$ holds, as asserted.

Concerning the left inclusion: One has that $\varphi(x) + \varphi'_x(v) \in H$ for all $x \in K$ and $v \in \sigma Q$ whenever σ is sufficiently small. We may then form

$$\bar{u} := \psi\big(\varphi(x) + (1 - \eta)\varphi'_x(v)\big) - x - (1 - \eta)v$$

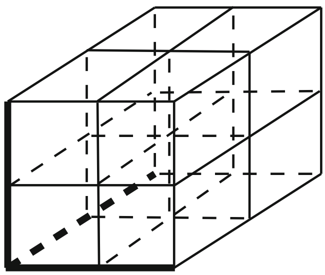
for all $x \in K$, $v \in \sigma Q$. By (10.4), $|\bar{u}| \leq \varepsilon\eta|v|$ if σ is sufficiently small. If ε is sufficiently small, it follows that $\bar{u} \in \eta\sigma Q$ and $(1 - \eta)v + \bar{u} \in \sigma Q$. According to the definition of $\bar{u}$ we obtain

$$\varphi(x) + (1 - \eta)\varphi'_x(v) = \varphi\big(x + (1 - \eta)v + \bar{u}\big) ,$$

and so $\varphi(x) + (1 - \eta)\varphi'_x(\sigma Q) \subset \varphi(x + \sigma Q)$.

□

Proof of the proposition. Let moreover $Q = [-c, c)$. We first determine the Lebesgue measure of $\varphi(z + Q)$, assuming that the topological closure K of $z + Q$ is contained in G. To this end we use that $z + Q$ can be partitioned, for any natural number n, into n^d disjoint rectangles $Q_{in} = x_{in} + n^{-1}Q$, $i = 1, \ldots, n^d$, with $x_{in} \in K$. The following figure illustrates the case $d = 3, n = 2$.



Since φ is bijective, the partition can be transferred to $\varphi(z+Q)$. Using the additivity, monotonicity, and translation invariance of the Lebesgue measure, we conclude from the lemma that for sufficiently large n

$$\sum_{i=1}^{n^d} \lambda^d\big((1 - \eta)\varphi'_{x_{in}}(n^{-1}Q)\big) \leq \lambda^d\big(\varphi(z + Q)\big) \leq \cdots$$

where we have omitted the upper estimate (with η instead of $-\eta$). We know the behaviour of the Lebesgue measure under linear mappings from Proposition 3.4, so it follows that

$$(1-\eta)^d \sum_{i=1}^{n^d} |\det \varphi'_{x_{in}}| \lambda^d(n^{-1}Q) \;\le\; \lambda^d\big(\varphi(z+Q)\big) \;\le\; (1+\eta)^d \cdots$$

or, written by means of an integral,

$$(1-\eta)^d \int \sum_{i=1}^{n^d} |\det \varphi'_{x_{in}}| 1_{Q_{in}}\, d\lambda^d \;\le\; \lambda^d\big(\varphi(z+Q)\big) \;\le\; (1+\eta)^d \cdots$$

Since $|\det \varphi'_x|$ is continuous, the integrands are uniformly bounded by a constant, and for $n \to \infty$ they converge to $|\det \varphi'_x| 1_{z+Q}$. By the dominated convergence theorem,

$$(1-\eta)^d \int |\det \varphi'_x| 1_{z+Q}\, dx \;\le\; \lambda^d\big(\varphi(z+Q)\big) \;\le\; (1+\eta)^d \cdots$$

and letting $\eta \to 0$ we finally obtain that

$$\lambda^d\big(\varphi(z+Q)\big) = \int_{z+Q} |\det \varphi'_x|\, dx\,.$$

This proves the formula for half-open rectangles. Consequently, it also holds for any finite disjoint union of such rectangles whose topological closure is contained in G. The system of all such unions forms a $\cap$-stable generator of the σ-algebra of all Borel sets $B \subset G$. In addition, it satisfies the assumptions of the uniqueness theorem, applied to the measures

$$\mu(B) := \lambda^d\big(\varphi(B)\big)\,, \quad \nu(B) := \int_B |\det \varphi'_x|\, dx$$

with $B \subset G$, because the open set G can be represented as a countable union of such rectangles. This yields the assertion. □

With the aid of the monotonicity principle (Proposition 2.8) we now obtain the following "substitution formula" for integration.

Corollary ***(Transformation Formula of Jacobi).*** *For any* C^1*-diffeomorphism* $\varphi : G \to H$ *and any nonnegative measurable function* $f : H \to \bar{\mathbb{R}}_+$ *it holds that*

$$\int_H f(y)\,dy = \int_G f(\varphi(x)) \cdot |\det \varphi'_x|\,dx\,.$$

Proof. For the Borel set $B' = \varphi(B)$, Proposition 10.1 can be rewritten as

$$\int_H 1_{B'}(y)\,dy = \int_G 1_{B'} \circ \varphi(x) \cdot |\det \varphi'_x|\,dx\,.$$

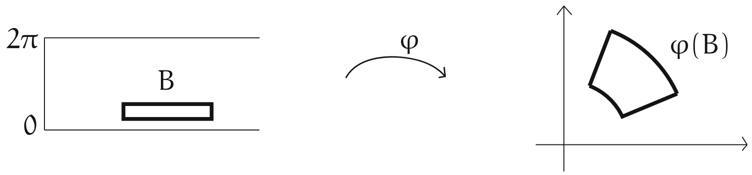
The assertion now follows from the monotonicity principle, Proposition 2.8. □

For integrable functions an analogous formula holds. The following example includes a well-known application.

Example (Polar coordinates)

The mapping

$$x = (r, \alpha) \ \mapsto \ y = (u, v) := (r\cos\alpha, r\sin\alpha)$$



defines a C^1-diffeomorphism from $G := (0, \infty) \times (0, 2\pi)$ to $H = \mathbb{R}^2 \setminus \{0\} \times \mathbb{R}_+$. The figure shows that the dilation is variable and equals r. Indeed, we obtain for the Jacobi determinant

$$\det \varphi'_x = \det \begin{pmatrix} \partial u/\partial r & \partial u/\partial \alpha \\ \partial v/\partial r & \partial v/\partial \alpha \end{pmatrix} = \det \begin{pmatrix} \cos\alpha & -r\sin\alpha \\ \sin\alpha & r\cos\alpha \end{pmatrix} = r\,.$$

One obtains an interesting application of the transformation formula for

$$f(y) = \exp(-|y|^2) = \exp(-u^2 - v^2)\,.$$

The formula yields

$$\int_{\mathbb{R}^2} \exp(-u^2)\exp(-v^2)\,du dv = \int_G \exp(-r^2) r\,d\alpha dr\,,$$

where we have extended H to $\mathbb{R}^2$ by the null set $\{0\} \times \mathbb{R}_+$. By Fubini's Theorem from Chap. 8 we may replace both two-dimensional Lebesgue integrals with double integrals in the respective variables, thus

$$\int_{-\infty}^{\infty} \exp(-u^2)\,du \int_{-\infty}^{\infty} \exp(-v^2)\,dv = \int_0^{2\pi} d\alpha \int_0^{\infty} \exp(-r^2) r\,dr = 2\pi \cdot \frac{1}{2}\,.$$

We obtain, for a second time, the formula

$$\int_{-\infty}^{\infty} \exp(-u^2)\,du = \sqrt{\pi}\,.$$

This argument goes back to Gauss.[2]

Exercises

10.1 Compute

$$\iint_B x^2 y^2\,dxdy\,,$$

where $B := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

[2] Carl Friedrich Gauss, 1777–1855, born in Braunschweig, active in Braunschweig and at the observatory in Göttingen. His contributions shape the whole of mathematics until the present time. For astronomy, physics and geodesy, too, he has lasting merits.

Construction of Measures

11

Let $\mathcal{A}$ be a σ-algebra on S with generator $\mathcal{E}$, and let

$$\pi : \mathcal{E} \to \bar{\mathbb{R}}_+$$

be a mapping which associates a nonnegative number $\pi(E)$ (or possibly the value ∞) to each element E of the generator. In this section we want to specify conditions which allow us to extend π to a measure μ on $\mathcal{A}$. More precisely, following Carathéodory we ask under which circumstances we may use for this purpose the mapping

$$\mu : \mathcal{A} \to \bar{\mathbb{R}}_+$$

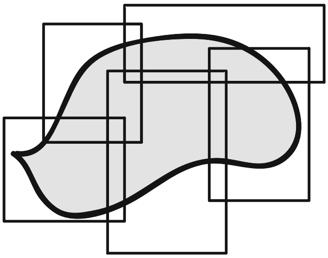
related to π, given by

$$\mu(A) := \inf \Big\{ \sum_{m \geq 1} \pi(E_m) : E_1, E_2, \ldots \in \mathcal{E},\ A \subset \bigcup_{m \geq 1} E_m \Big\} .$$

As usual we set $\inf \varnothing := \infty$. (While we follow up our discussion concerning the regularity of measures in Chap. 7, we do not need its results in the following.)

Thus, the idea is to obtain the measure of A by outer approximation, covering A with finitely or infinitely many elements $E_1, E_2, \ldots$ from $\mathcal{E}$

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_11



the sum of their measures becoming as small as possible. The question is under which conditions this procedure works. We will also discuss some applications.

As a preparation, we first present a general method to construct a measure from an outer measure, dating back to Carathéodory. This method has a large scope of application and in particular yields, for example, the Hausdorff measures discussed at the end of this chapter.

Outer Measures

Definition

A mapping

$$\eta : \mathcal{P}(S) \to \bar{\mathbb{R}}_+$$

on the power set $\mathcal{P}(S)$ of S is called an *outer measure* if the following holds:

(i) $\eta(\varnothing) = 0$,
(ii) σ-subadditivity: $\eta(A) \leq \sum_{n\geq 1} \eta(A_n)$ for all $A, A_1, A_2, \ldots \subset S$ which satisfy $A \subset \bigcup_{n\geq 1} A_n$.

A subset $A \subset S$ is called η-*measurable* if for every $C \subset S$

$$\eta(C \cap A) + \eta(C \cap A^c) = \eta(C)$$

holds.

In particular, the σ-subadditivity entails the property of

(iii) Monotonicity: $\eta(A) \leq \eta(A')$, whenever $A \subset A'$.

η-measurability of A means that one may separate η into two parts on A and on A^c, from which one may get back η by addition. For the η-measurability of A it suffices that $\eta(C\cap A)+\eta(C\cap A^c) \le \eta(C)$ holds for all $C \subset \mathcal{S}$, because subadditivity yields the reverse inequality.

The following result holds.

Proposition 11.1 (Carathéodory). *Let η be an outer measure on* S. *Then the system $\mathcal{A}_\eta$ of all η-measurable sets is a σ-algebra, and the restriction of η to $\mathcal{A}_\eta$ is a measure.*

Proof. Immediately obvious are the properties

$$S \in \mathcal{A}_\eta \quad \text{and} \quad A \in \mathcal{A}_\eta \;\Rightarrow\; A^c \in \mathcal{A}_\eta \ .$$

Let $A_1, A_2 \in \mathcal{A}_\eta$. Repeated application of the defining property of η-measurable subsets yields

$$\begin{aligned}\eta(C) &= \eta(C \cap A_1) + \eta(C \cap A_1^c) \qquad (*)\\ &= \eta(C \cap A_1) + \eta(C \cap A_1^c \cap A_2) + \eta(C \cap A_1^c \cap A_2^c)\\ &= \eta(C \cap (A_1 \cup A_2) \cap A_1) + \eta(C \cap (A_1 \cup A_2) \cap A_1^c)\\ &\qquad + \eta(C \cap (A_1 \cup A_2)^c)\\ &= \eta(C \cap (A_1 \cup A_2)) + \eta(C \cap (A_1 \cup A_2)^c) \ .\end{aligned}$$

It follows that $A_1 \cup A_2 \in \mathcal{A}_\eta$ as well as $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{A}_\eta$. If A_1 and A_2 are disjoint, we obtain from row $(*)$, replacing C by $C \cap (A_1 \cup A_2)$, the additivity property

$$\eta(C \cap (A_1 \cup A_2)) = \eta(C \cap A_1) + \eta(C \cap A_2) \ .$$

Let moreover $A_1, A_2, \ldots \in \mathcal{A}_\eta$ be pairwise disjoint. Using the additivity just proved, as well as the monotonicity of η, we obtain that for any natural number r

$$\begin{aligned}\eta(C) &= \eta\Big(C \cap \bigcup_{n=1}^{r} A_n\Big) + \eta\Big(C \cap \Big(\bigcup_{n=1}^{r} A_n\Big)^c\Big)\\ &\ge \sum_{n=1}^{r} \eta\Big(C \cap A_n\Big) + \eta\Big(C \cap \Big(\bigcup_{n\ge1} A_n\Big)^c\Big) \ .\end{aligned}$$

Letting $r \to \infty$ and using σ-subadditivity we get

$$\begin{aligned} \eta(C) &\geq \sum_{n\geq 1} \eta(C \cap A_n) + \eta\Big(C \cap \Big(\bigcup_{n\geq 1} A_n\Big)^c\Big) \qquad (**) \\ &\geq \eta\Big(C \cap \bigcup_{n\geq 1} A_n\Big) + \eta\Big(C \cap \Big(\bigcup_{n\geq 1} A_n\Big)^c\Big) \\ &\geq \eta(C)\ . \end{aligned}$$

Thus, equality holds everywhere above, and $\bigcup_{n\geq 1} A_n \in \mathcal{A}_\eta$ follows. The particular choice $C = \bigcup_{n\geq 1} A_n$ in row $(**)$ yields

$$\eta\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \eta(A_n)\ ,$$

that is, η is σ-additive on $\mathcal{A}_\eta$. Finally, arbitrary countable unions can be reduced to disjoint unions according to

$$\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} A_n \cap A_1^c \cap \cdots \cap A_{n-1}^c\ ,$$

so that $\mathcal{A}_\eta$ indeed is a σ-algebra. □

Extension to a Measure

We utilize outer measures to prove the following proposition.

Proposition 11.2 (Extension). *Let $\mathcal{E}$ generate the σ-algebra $\mathcal{A}$ on S, and let $\pi : \mathcal{E} \to \bar{\mathbb{R}}_+$ be a mapping. Then*

$$\mu(A) := \inf\Big\{\sum_{m\geq 1} \pi(E_m) : E_1, E_2, \ldots \in \mathcal{E},\ A \subset \bigcup_{m\geq 1} E_m\Big\}\ , \qquad A \in \mathcal{A}\ ,$$

defines a measure μ on $\mathcal{A}$ coinciding with π on $\mathcal{E}$, if and only if the conditions

(i) $\mu(\emptyset) = 0$,
(ii) $\mu(E) = \pi(E)$ *for all* $E \in \mathcal{E}$,
(iii) $\mu(E' \cap E) + \mu(E' \cap E^c) \leq \pi(E')$ *for all* $E, E' \in \mathcal{E}$

are satisfied.

Since $\mu(E) \le \pi(E)$ always holds, (ii) can be replaced with $\mu(E) \ge \pi(E)$. The verification of this seemingly innocuous condition typically requires a substantial effort. According to the definition of μ it is equivalent to the condition

(ii') $\pi(E) \le \sum_{m\ge 1} \pi(E_m)$ for $E, E_1, E_2, \ldots \in \mathcal{E}$ and $E \subset \bigcup_{m\ge 1} E_m$.

We will see how, in order to verify it, one replaces the infinite covering of E with other, more easily tractable, finite coverings. Reasonings like that, based on compactness arguments, go back to Borel, who established the topological concept of compactness in mathematics.

Proof. The conditions are obviously necessary. To prove their sufficiency we extend μ to the whole power set by

$$\eta(A) := \inf\Big\{\sum_{m\ge 1} \pi(E_m) : E_1, E_2, \ldots \in \mathcal{E}, A \subset \bigcup_{m\ge 1} E_m\Big\} \quad \text{for all } A \subset S .$$

η is σ-subadditive: Let $A, A_1, A_2, \ldots \subset S$ be such that $A \subset \bigcup_{n\ge 1} A_n$ holds. By definition of η, for every $\varepsilon > 0$ there exist elements $E_{1n}, E_{2n}, \ldots$ of $\mathcal{E}$ such that $A_n \subset \bigcup_{m\ge 1} E_{mn}$ and

$$\sum_{m\ge 1} \pi(E_{mn}) \le \eta(A_n) + \varepsilon 2^{-n} .$$

It follows that $A \subset \bigcup_{m,n\ge 1} E_{mn}$ and

$$\eta(A) \le \sum_{m,n\ge 1} \pi(E_{mn}) \le \sum_{n\ge 1}(\eta(A_n) + \varepsilon 2^{-n}) \le \sum_{n\ge 1} \eta(A_n) + \varepsilon .$$

Letting $\varepsilon \to 0$ we obtain the σ-subadditivity. By (i) we moreover have $\eta(\emptyset) = 0$, thus η is an outer measure.

We show next that each $E \in \mathcal{E}$ is a η-measurable set. Let $C \subset S$ and $E_1, E_2, \ldots \in \mathcal{E}$ satisfying $C \subset \bigcup_{m\ge 1} E_m$. Since η is σ-subadditive,

$$\eta(C) \le \eta(C \cap E) + \eta(C \cap E^c) \le \sum_{m\ge 1} \eta(E_m \cap E) + \sum_{m\ge 1} \eta(E_m \cap E^c) ,$$

and due to (iii)

$$\eta(C) \le \eta(C \cap E) + \eta(C \cap E^c) \le \sum_{m\ge 1} \pi(E_m) .$$

According to the definition of η, we may choose, for any given $\varepsilon > 0$, the sets $E_1, E_2, \ldots$ in such a way that $\sum_{m\geq 1} \pi(E_m) \leq \eta(C) + \varepsilon$ holds. It follows that

$$\eta(C) \leq \eta(C \cap E) + \eta(C \cap E^c) \leq \eta(C) + \varepsilon\,.$$

Letting $\varepsilon \to 0$ we see that E is η-measurable.

We can now apply the preceding proposition. Since $\mathcal{E}$ generates $\mathcal{A}$, it follows at first that $\mathcal{A} \subset \mathcal{A}_\eta$ and secondly that μ is a measure. Due to condition (ii), μ coincides with π on $\mathcal{E}$ as claimed. □

By its very definition, the measure obtained in the extension theorem is outer regular w.r.t. $\mathcal{E}$. The theorem has important applications.

Example (Locally finite measures on $\mathbb{R}$)

We consider measures on $\mathbb{R}$ which are finite on bounded sets. Such measures μ are, as a consequence of the uniqueness theorem, uniquely determined by the values

$$\mu\big((a, b]\big)\,, \quad -\infty < a \leq b < \infty\,.$$

One always can exhibit a "primitive" $F : \mathbb{R} \to \mathbb{R}$ such that

$$\mu\big((a, b]\big) = F(b) - F(a)$$

holds, e. g., $F(a) := \mu\big((0, a]\big)$, resp. $-\mu\big((a, 0]\big)$, depending on whether $a \geq 0$ or $a < 0$. In addition, F (like primitives in calculus) is uniquely determined by μ up to a constant. F is obviously monotone and, by virtue of the σ-continuity of μ, right-continuous.

Here we want to show that, conversely, for every monotone right-continuous function F there exists a measure μ such that the stated connection stands. For this purpose, on the generator

$$\mathcal{E} := \{(a, b] : -\infty < a \leq b < \infty\}$$

of the Borel σ-algebra in $\mathbb{R}$ we consider the functional $\pi : \mathcal{E} \to \mathbb{R}$ given by

$$\pi\big((a, b]\big) := F(b) - F(a)\,.$$

We want to show that the conditions of the extension theorem are satisfied.

Obviously $\pi(\emptyset) = 0$ holds, thus (i) is satisfied. Moreover, for any $E' = (a', b']$ and $E \in \mathcal{E}$ there exist numbers $a' \leq a \leq b \leq b'$ such that

$$E' \cap E = (a, b]\,, \quad E' \cap E^c = (a', a] \cup (b, b']\,.$$

Consequently, $\mu(E' \cap E) \leq \pi((a, b])$ and $\mu(E' \cap E^c) \leq \pi((a', a]) + \pi((b, b'])$, whence

$$\mu(E' \cap E) + \mu(E' \cap E^c) \leq F(b') - F(a') = \pi(E')\,.$$

Therefore (iii) is satisfied.

Finally, let $(a, b] \subset \bigcup_{m\geq 1}(a_m, b_m]$. As mentioned before, in order to prove (ii') we will pass from the countable covering to suitable finite coverings: Since F is right-continuous, for any given

$\varepsilon > 0$ there exist numbers $\varepsilon_m > 0$ such that $F(b_m + \varepsilon_m) \leq F(b_m) + \varepsilon 2^{-m}$. It follows that $[a + \varepsilon, b] \subset \bigcup_{m \geq 1} (a_m, b_m + \varepsilon_m)$. Thus, we exhibited an open covering of a compact set which accordingly contains a finite subcovering. We thus have $(a + \varepsilon, b] \subset \bigcup_{m=1}^{n} (a_m, b_m + \varepsilon_m]$ for a sufficiently large natural number n, and consequently

$$F(b) - F(a + \varepsilon) \leq \sum_{m=1}^{n} \big(F(b_m + \varepsilon_m) - F(a_m)\big) \leq \sum_{m=1}^{n} \big(F(b_m) - F(a_m)\big) + \varepsilon \; .$$

Letting first $n \to \infty$ and then $\varepsilon \to 0$ we obtain that

$$\pi\big((a, b]\big) \leq \sum_{m \geq 1} \pi\big((a_m, b_m]\big) \; .$$

Therefore, (ii') is satisfied.

By virtue of the extension theorem, there exists a measure μ on the Borel σ-algebra satisfying

$$\mu\big((a, b]\big) = F(b) - F(a) \; ,$$

as claimed.

Example (Lebesgue measure)

In the particular case $F(a) = a$, as in the foregoing example, one obtains the 1-dimensional Lebesgue measure. The d-dimensional Lebesgue measure can be constructed analogously, using d-dimensional instead of 1-dimensional intervals, or alternatively as a product measure from the 1-dimensional Lebesgue measure.

Outer Regularity*

Now we can prove the proposition concerning outer regularity in Chap. 7. Let us repeat its assertion (with a change in notation).

Let $\mathcal{E}$ be a $\cap$-stable generator of the σ-algebra $\mathcal{A}$ on S with $\varnothing \in \mathcal{E}$. Let ν be a measure on $\mathcal{A}$ such that there exist $E_1, E_2, \ldots \in \mathcal{E}$ with $E_m \uparrow S$ and $\nu(E_m) < \infty$ for all $m \geq 1$. Set

$$\mu(A) := \inf\Big\{ \sum_{m \geq 1} \nu(E_m) : E_1, E_2, \ldots \in \mathcal{E}, A \subset \bigcup_{m \geq 1} E_m \Big\} \; , \quad A \in \mathcal{A} \; .$$

If

$$\mu(E' \setminus E) = \nu(E' \setminus E) \quad \text{for all } E, E' \in \mathcal{E} \text{ with } E \subset E' \; ,$$

then ν is outer regular w.r.t. $\mathcal{E}$, that is, $\nu(A) = \mu(A)$ for all $A \in \mathcal{A}$.

Proof. We show that the conditions of the extension theorem hold with $\pi := \nu|\mathcal{E}$. Condition (ii') is satisfied automatically because ν is a measure. Thus (ii) holds and, as $\varnothing \in \mathcal{E}$, also (i).

Concerning (iii): Let $E, E' \in \mathcal{E}$. As $\mathcal{E}$ is $\cap$-stable, $\nu(E' \cap E) = \mu(E' \cap E)$ holds. By assumption we moreover have $\nu(E' \cap E^c) = \nu(E' \setminus E \cap E') = \mu(E' \setminus E \cap E') = \mu(E' \cap E^c)$. Since ν is additive,

$$\mu(E' \cap E) + \mu(E' \cap E^c) = \nu(E') = \pi(E')$$

follows. Thus (iii) is verified.

By the extension theorem, μ is a measure which coincides with ν on $\mathcal{E}$. Applying the uniqueness theorem for measures we obtain that $\mu = \nu$. □

The Riesz Representation Theorem

Any measure μ induces a functional $f \mapsto \ell(f) := \int f \, d\mu$ on the space of μ-integrable functions. This functional is linear and positive, that is,

$$\ell(\alpha f + \beta g) = \alpha\ell(f) + \beta\ell(g) \text{ for all } \alpha, \beta \in \mathbb{R} , \quad \ell(f) \geq 0 \text{ for all } f \geq 0 .$$

If $f \leq g$, it follows that $\ell(f) \leq \ell(g)$, since $\ell(g) - \ell(f) = \ell(g - f) \geq 0$.

Conversely, we may ask which positive linear functionals can be represented as integrals. We consider the simplest, yet most important case. By $C(S)$ we denote the linear space of all real-valued continuous functions on a metric space S.

Proposition 11.3 (Riesz Representation Theorem). *Let S be a compact metric space endowed with the Borel σ-algebra $\mathcal{B}$, and let $\ell : C(S) \to \mathbb{R}$ be a positive linear functional. Then there exists a unique measure μ on $(S, \mathcal{B})$ satisfying*

$$\ell(f) = \int f \, d\mu$$

for every $f \in C(S)$.

We obtain such a measure with the aid of the extension Proposition 11.2. As generator $\mathcal{E}$ of $\mathcal{B}$ we choose the system $\mathcal{O}$ of all open subsets of S. We define a set function $\pi : \mathcal{O} \to \mathbb{R}_+$ by

$$\pi(O) = \sup_{0 \leq f \leq 1_O} \ell(f) .$$

Since $0 \leq 1_O \leq 1$, we have $0 \leq \pi(O) \leq \ell(1)$. We directly conclude that $\pi(\varnothing) = 0$, $\pi(O) \leq \pi(O')$ whenever $O \subset O'$, as well as $\ell(f) \leq \pi(O) \leq \ell(g)$ whenever we have

$f \le 1_O \le g$. (We remark that the definition $\pi(O) = \ell(1_O)$ is not possible as 1_O is, in general, not continuous.)

We define

$$\mu(A) = \inf\Big\{\sum_{m\ge 1} \pi(O_m) : O_1, O_2, \dots \in \mathcal{O}, A \subset \bigcup_{m\ge 1} O_m\Big\}, \quad A \in \mathcal{B}.$$

The extension Proposition 11.2 for measures says that μ defines a measure on $\mathcal{B}$ if the conditions

(i) $\pi(\varnothing) = 0$,
(ii) $\mu(O) = \pi(O)$ for all $O \in \mathcal{O}$,
(iii) $\mu(O' \cap O) + \mu(O' \cap O^c) \le \pi(O')$ for all $O, O' \in \mathcal{O}$,

are satisfied, where

(ii′) $\pi(O) \le \sum_{m\ge 1} \pi(O_m)$ for all $O, O_1, O_2, \ldots \in \mathcal{O}$ with $O \subset \bigcup_{m\ge 1} O_m$,

is equivalent to (ii).

As a prerequisite we provide a connection between monotonicity and uniform convergence in $C(S)$.

Lemma (Dini's[1] Theorem). *Let* $f_1, f_2, \ldots$ *be a sequence in* $C(S)$ *with* $f_n \uparrow f$ *and* $f \in C(S)$, S *being a compact metric space. Then* f_n *converges uniformly to* f.

Proof. Let $\varepsilon > 0$. Given $x \in S$, we choose n_x such that $|f(x) - f_{n_x}(x)| < \varepsilon$. Due to continuity there exists an open neighbourhood O_x of x such that $|f - f_{n_x}| < \varepsilon$ holds on O_x. By virtue of compactness, S can be covered by finitely many such O_x, say for points x_j, $1 \le j \le m$. It follows that $\|f - f_n\|_\infty < \varepsilon$ for any $n \ge \max_j n_{x_j}$. □

We return to the task of proving properties (i) to (iii) and consider, for any $O \in \mathcal{O}$, the functions

$$\varphi_{n,O}(x) = \min\big(1, n d(x, O^c)\big).$$

They are continuous and satisfy $0 \le \varphi_{1,O} \le \varphi_{2,O} \le \cdots$, as well as

$$1_O = \sup_{n\ge 1} \varphi_{n,O}, \qquad \pi(O) = \sup_{n\ge 1} \ell(\varphi_{n,O}).$$

[1] ULISSE DINI, 1845–1918, born in Pisa, active in Pisa. He did research in real analysis.

The latter is a consequence of Dini's Theorem, since for any $f \le 1_O$ and $f_n = \min(f, \varphi_{n,O})$ one has that $f_n \uparrow f$, and thus

$$\ell(f) = \sup_{n\ge1} \ell(f_n) \le \sup_{n\ge1} \ell(\varphi_{n,O}) \le \pi(O)\,,$$

which after passing to the supremum w.r.t. f implies the asserted equality.

Lemma. *The set function* π *satisfies* (ii′) *and consequently* (ii).

Proof. Let $O, O_1, O_2, \ldots \in \mathcal{O}$ satisfying $O \subset \bigcup_{m\ge1} O_m$ be given, let $f \le 1_O$. We set $g_n = \sum_{m=1}^{n} \varphi_{n,O_m}$ and $f_n = \min(f, g_n)$. Then

$$\ell(f_n) \le \ell(g_n) = \sum_{m=1}^{n} \ell(\varphi_{n,O_m}) \le \sum_{m=1}^{n} \pi(O_m)$$

holds. Since $f \le 1_O \le \sup_{n\ge1} g_n$ we have $f_n \uparrow f$, thus $\ell(f) = \sup_{n\ge1} \ell(f_n) \le \sum_{m\ge1} \pi(O_m)$ by Dini's Theorem, and the assertion follows after passing to the supremum w.r.t. f. □

Lemma. *The set function* π *satisfies* (iii).

In this proof, for any $O \in \mathcal{O}$, instead of $\varphi_{n,O}$ we use

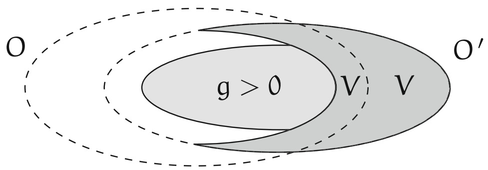
$$\psi_{n,O}(x) := \min\left(1, (nd(x, O^c) - 1)^+\right).$$

$\psi_{n,O}$ has the same properties as those we just derived for $\varphi_{n,O}$. In addition, $d(x, O^c) > 1/n$ whenever $\psi_{n,O}(x) > 0$.

Proof. Let O, O' be open sets. We set $g := \psi_{n,O'\cap O}$ and

$$V := \{x \in O' : d(x, O^c) < 1/n\}\,.$$

One has that $\{g > 0\} \cap V = \varnothing$.



Let $\varepsilon > 0$. We choose n large enough such that $\pi(O' \cap O) \le \ell(g) + \varepsilon$. Since V is open, there exists an $h \le 1_V$ satisfying $\pi(V) \le \ell(h) + \varepsilon$. We have $0 \le g + h \le 1_{O'}$, as $V \subset O'$ and $g(x) = 0$ for any $x \in V$. Since moreover $O' \cap O^c \subset V$, it follows that

$$\begin{aligned}\mu(O' \cap O) + \mu(O' \cap O^c) &\le \pi(O' \cap O) + \pi(V) \le \ell(g + h) + 2\varepsilon \\ &\le \pi(O') + 2\varepsilon\,.\end{aligned}$$

Passing to the limit $\varepsilon \to 0$ yields the assertion. □

Proof of Riesz' theorem. The preceding lemmata show that the assumptions of the extension Proposition 11.2 are satisfied, implying that μ is a measure on $\mathcal{B}$, and that $\mu(S) = \ell(1) < \infty$. It remains to show that $\ell(f) = \int f\,d\mu$ holds for all $f \in C(S)$. Let $f \ge 0$ be continuous. For any $n \ge 1$ and any $k \ge 0$ we set

$$f_{kn} = \min\Big(\frac{1}{n}, (f - \frac{k}{n})^+\Big)\,.$$

The functions f_{kn} are continuous, and for every n one has $f = \sum_{k\ge 0} f_{kn}$, where at most finitely many summands are nonzero, as well as

$$\frac{1}{n} 1_{\{f>(k+1)/n\}} \le f_{kn} \le \frac{1}{n} 1_{\{f>k/n\}}\,.$$

According to the definition of π,

$$\frac{1}{n}\pi(\{f > (k+1)/n\}) \le \ell(f_{kn}) \le \frac{1}{n}\pi(\{f > k/n\})\,,$$

thus

$$\begin{aligned}\ell(f) &= \sum_{k\ge 0}\ell(f_{kn}) \le \sum_{k\ge 0}\frac{1}{n}\mu(\{k/n < f\}) = \sum_{k\ge 0}\frac{k+1}{n}\mu(\{k/n < f \le (k+1)/n\}) \\ &\le \int f\,d\mu + \frac{1}{n}\mu(S)\,.\end{aligned}$$

Letting $n \to \infty$ we get $\ell(f) \le \int f\,d\mu$. The reverse inequality results in an analogous manner, and so $\ell(f) = \int f\,d\mu$ for any $f \ge 0$ and, via a decomposition in positive and negative part, for arbitrary $f \in C(S)$, as it was claimed. The uniqueness of μ was already proved earlier in Chap. 7. □

We remark that the measure μ, constructed in the Riesz representation theorem, is regular by Proposition 7.6.

Extension of Measures on Infinite Products*

The following result, which goes back to Kolmogorov,[2] is of interest in probability theory. The question is this: Let there be given finite measures μ_d on the Borel σ-algebra $\mathcal{B}^d$ of $\mathbb{R}^d$, $d \geq 1$. Under which conditions does there exist a measure μ on the product space $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ which extends the measures μ_d in the sense that

$$\mu(B \times \mathbb{R}^\infty) = \mu_d(B) , \quad B \in \mathcal{B}^d$$

holds? Such a μ is called the *projective limit* of the measures μ_d. The measures μ_d obviously have to be related to each other in the following manner.

Definition

A sequence μ_d, $d \geq 1$, of finite measures on $\mathbb{R}^d$ is called *consistent*, if

$$\mu_{d+1}(B \times \mathbb{R}) = \mu_d(B)$$

for every $d \geq 1$ and every Borel set $B \in \mathcal{B}^d$.

Example (Product measures)

If $\mu_{d+1} = \mu_d \otimes \nu_{d+1}$ holds for some probability measures $\nu_2, \nu_3, \ldots$, then $\mu_1, \mu_2, \ldots$ are consistent measures.

Proposition 11.4 (Kolmogorov's Theorem). *Every consistent sequence* $\mu_1, \mu_2, \ldots$ *of finite measures possesses a unique projective limit* μ.

Proof. By $\mathcal{E}$ we denote the system of all sets $O \times \mathbb{R}^\infty \subset \mathbb{R}^\infty$, where O is an open subset of some $\mathbb{R}^d$ with $d = 1, 2, \ldots$. We define the set function $\pi : \mathcal{E} \to \mathbb{R}_+$ by

$$\pi(O \times \mathbb{R}^\infty) := \mu_d(O) .$$

Here we have to take into account that every $E \in \mathcal{E}$ allows different representations, namely $E = O \times \mathbb{R}^\infty$ can also be written as $E = O' \times \mathbb{R}^\infty$ with $O' = O \times \mathbb{R}^e$, $e \geq 1$. Nevertheless π is well defined due to the consistency condition.

$\mathcal{E}$ generates the product σ-algebra $\mathcal{B}^\infty$ on $\mathbb{R}^\infty$. We define μ as in the measure extension theorem and therefore have to verify the conditions of the latter.

Concerning condition (iii): For any $E, E' \in \mathcal{E}$ there exist a (common!) $d \geq 1$ and $O, O' \in \mathcal{B}^d$ such that $E = O \times \mathbb{R}^\infty$, $E' = O' \times \mathbb{R}^\infty$. Moreover, we consider the open

[2] Andrei N. Kolmogorov, 1903–1987, born in Tambov, active in Moscow. He made seminal contributions to probability theory, topology, dynamical systems, mechanics, and turbulent flows.

sets $O_n := \{x \in \mathbb{R}^d : |x-y| < 1/n \text{ for some } y \in O^c\}$, the open $1/n$-neighbourhoods of the closed set O^c. Setting $E_n = O_n \times \mathbb{R}^\infty$ we get for every $n \geq 1$

$$\mu(E' \cap E) + \mu(E' \cap E^c) \leq \pi(E' \cap E) + \pi(E' \cap E_n) = \mu_d(O' \cap O) + \mu_d(O' \cap O_n) ,$$

and passing to the limit $n \to \infty$ we obtain, by virtue of σ-continuity,

$$\mu(E' \cap E) + \mu(E' \cap E^c) \leq \mu_d(O' \cap O) + \mu_d(O' \cap O^c) = \mu_d(O') = \pi(E') .$$

This yields (iii). Condition (i) follows from (ii) since $\varnothing \in \mathcal{E}$.

It remains to prove (ii′): Let $E = O \times \mathbb{R}^\infty$, with $O \subset \mathbb{R}^d$ open. As we already did previously, we will pass from a countable covering of E to suitable finite coverings. To this end we choose $\varepsilon > 0$ and, according to Proposition 7.6, for every $n \geq 1$ a compact set $K_n \subset O \times \mathbb{R}^n$ such that $\pi(E) = \mu_{d+n}(O \times \mathbb{R}^n) < \mu_{d+n}(K_n) + \varepsilon$.

Let now $E \subset \bigcup_{m\geq 1} E_m$ satisfying $E_m \in \mathcal{E}$ and $E_m = O_m \times \mathbb{R}^\infty$. We want to show that there exists an $n \geq 1$ such that

$$K_n \times \mathbb{R}^\infty \subset \bigcup_{m=1}^{n} E_m .$$

Contrarily, assume that there exist $x_1, x_2, \ldots$ in $\mathbb{R}^\infty$ satisfying $x_n \in K_n \times \mathbb{R}^\infty$ and $x_n \notin \bigcup_{m=1}^n E_m$ for all $n \geq 1$. Then one may pass to a subsequence converging componentwise, by the following scheme: As K_1 is compact, there exists a subsequence $x_{i,1} \in \mathbb{R}^\infty$, $i \geq 1$, whose first $d+1$ components converge. As K_2 is compact, we find a subsubsequence $x_{i,2}$, $i \geq 1$, for which also the $(d+2)$-th component converges. One continues as follows: In the kth subsubsequence $x_{i,k}$, $i \geq 1$, the first $d+k$ components converge. Following Cantor, we finally pass to the diagonal sequence $x_{i,i} \in \mathbb{R}^\infty$, $i \geq 1$, which eventually traverses every subsequence and for which therefore all components converge to some limit $y = (y_1, y_2, \ldots)$. It follows that $y \in K_1 \times \mathbb{R}^\infty \subset E \subset \bigcup_{m\geq 1} E_m$ and thus $y \in E_j$ for some $j \geq 1$. Since O_j is open we conclude that $x_{i,i} \in E_j$ whenever i is sufficiently large. As the diagonal sequence is a subsequence of the original sequence x_n, $n \geq 1$, there exists an $n \geq j$ such that $x_n \in \bigcup_{m=1}^n E_m$. This is a contradiction.

Therefore, there exists an $n \geq 1$ such that the inclusion stated above holds. In other words, there exist a $k \geq n + d$ and open sets $O_m \in \mathbb{R}^k$, $m \leq n$ such that $E_m = O_m \times \mathbb{R}^\infty$ and $K_n \times \mathbb{R}^{k-n-d} \subset \bigcup_{m=1}^n O_m$. Due to the subadditivity of μ_k it follows that

$$\pi(E) - \varepsilon \leq \mu_{d+n}(K_n) = \mu_k(K_n \times \mathbb{R}^{k-n-d}) \leq \sum_{m=1}^{n} \mu_k(O_m) ,$$

therefore $\pi(E) \leq \sum_{m=1}^n \pi(E_m) + \varepsilon$. Passing to the limit $n \to \infty$ and then $\varepsilon \to 0$ we obtain (ii′).

In this manner the extension theorem yields a measure μ satisfying $\mu(O\times\mathbb{R}^\infty) = \mu_d(O)$ for all open $O \subset \mathbb{R}^d$. Using the uniqueness theorem $\mu(B\times\mathbb{R}^\infty) = \mu_d(B)$ for all Borel sets $B \subset \mathbb{R}^d$. Therefore, μ is the projective limit of the measures μ_d, $d \geq 1$. As $\mathcal{E}$ is a $\cap$-stable generator of $\mathcal{B}^\infty$, the projective limit is uniquely determined. □

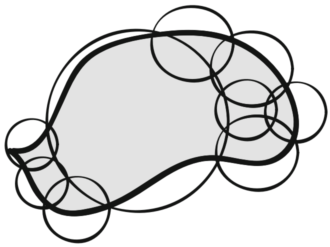
The compactness argument in the proof above may also be based on Tikhonov's Theorem which states that infinite Cartesian products of compact sets are themselves compact. This would shorten the proof.

Kolmogorov's theorem can be generalized in several directions. The space $\mathbb{R}$ may be replaced by spaces in which open subsets can be approximated from the interior by compact sets, at least in measure. This works for all complete separable metric spaces (Ulam's Theorem). The result may also be transferred to uncountable products without a major effort.

Hausdorff Measures*

The Lebesgue measure is not the only translation invariant measure on the Borel sets of $\mathbb{R}^d$. As a conclusion of this chapter we want to present a whole family of translation invariant measures. Only if the unit cube has finite measure, one is dealing (except for a normalization factor) with the Lebesgue measure.

A basic idea is to cover a subset A of $\mathbb{R}^d$ by balls and other sets of bounded diameter



and to obtain a number measuring A from their count and their diameter. There are different possibilities to do so, one may assign a positive measure also for "sparse" sets with Lebesgue measure 0. It appears to be natural to cover A with sets of very small diameter only – we will see that there are sound mathematical reasons for this. Our approach makes use of outer measures η_s which depend on a given parameter $s > 0$.

We define the *diameter* of $A \subset \mathbb{R}^d$ as

$$d(A) := \sup\left\{|x-y| : x, y \in A\right\} .$$

In a first step we prescribe (besides s) a $\delta > 0$ and set

$$\eta_{s,\delta}(A) := \inf\Big\{\sum_{m\geq 1} d(A_m)^s : A \subset \bigcup_{m\geq 1} A_m \ ,\ d(A_m) \leq \delta\Big\}\ , \quad A \subset \mathbb{R}^d\ .$$

For covering we thus use arbitrary sets of diameter at most δ.

The set function $\eta_{s,\delta}$ is an outer measure; this is proved as for the extension theorem above. However, in general one does not know what are the corresponding measurable sets. Therefore, in a second step one passes to

$$\eta_s(A) := \sup_{\delta>0} \eta_{s,\delta}(A)\ , \quad A \subset \mathbb{R}^d\ .$$

This means that henceforth we consider only small δ, since $\eta_{s,\delta}(A)$ increases monotonically as δ decreases. Obviously η_s is translation invariant.

η_s, too, is an outer measure: since $\eta_{s,\delta}(\varnothing) = 0$ we also have $\eta_s(\varnothing) = 0$, and from $\eta_{s,\delta}\big(\bigcup_{n\geq 1} A_n\big) \leq \sum_{n\geq 1} \eta_{s,\delta}(A_n) \leq \sum_{n\geq 1} \eta_s(A_n)$ follows $\eta_s\big(\bigcup_{n\geq 1} A_n\big) \leq \sum_{n\geq 1} \eta_s(A_n)$.

For η_s an additional property comes into play: Let

$$a(A', A'') := \inf\big\{|x-y| : x \in A'\ ,\ y \in A''\big\}\ ,$$

denote the *distance* between two subsets A', A'' of $\mathbb{R}^d$ (using the convention $\inf\varnothing = \infty$, the distance to the empty set equals ∞). Following Carathéodory, an outer measure η on $\mathbb{R}^d$ is called *metric* if it satisfies the condition

$$a(A', A'') > 0 \quad \Rightarrow \quad \eta(A' \cup A'') = \eta(A') + \eta(A'')\ .$$

The outer measures η_s are metric. Indeed, let $A' \cup A'' \subset \bigcup_{m\geq 1} A_m$ where $d(A_m) \leq \delta$. If $\delta < a(A, A')/2$, then each A_m intersects at most one of the sets A', A''. Therefore we may partition the sequence A_m into two subsequences A'_m, A''_m, $m \geq 1$, such that $A' \subset \bigcup_{m\geq 1} A'_m$ and $A'' \subset \bigcup_{m\geq 1} A''_m$. It follows that $\sum_{m\geq 1} d(A_m)^s \geq \eta_{s,\delta}(A') + \eta_{s,\delta}(A'')$, therefore $\eta_{s,\delta}(A' \cup A'') \geq \eta_{s,\delta}(A') + \eta_{s,\delta}(A'')$ and, letting $\delta \to 0$, finally $\eta_s(A' \cup A'') \geq \eta_s(A') + \eta_s(A'')$. The reverse inequality holds because η_s is an outer measure.

The importance of metric outer measures stems from the following characterization.

Proposition 11.5. *An outer measure η on $\mathbb{R}^d$ is metric if and only if all Borel sets are η-measurable.*

Proof. First assume that all Borel sets are η-measurable. If $a(A', A'') > 0$ holds for sets A', A'', then $O := \{y \in \mathbb{R}^d : |y - x| < a(A', A'')$ for some $x \in A'\}$ is an open set. Its η-measurability implies that

$$\eta(A' \cup A'') = \eta\big((A' \cup A'') \cap O\big) + \eta\big((A' \cup A'') \cap O^c\big) .$$

Moreover, we have $A' \subset O$, $A'' \subset O^c$, and we obtain $\eta(A' \cup A'') = \eta(A') + \eta(A'')$. Therefore η is a metric outer measure.

Conversely, let η be metric. We show that in this case every closed set $A \subset \mathbb{R}^d$ is η-measurable, so that $\eta(C) \geq \eta(C \cap A) + \eta(C \cap A^c)$ holds for all $C \subset \mathbb{R}^d$. Without loss of generality, we may assume that $\eta(C) < \infty$. To prove the claim we construct sets $D_1 \subset D_2 \subset \cdots \subset C \cap A^c$ such that $a(C \cap A, D_n) > 0$ and $\eta(D_n) \to \eta(C \cap A^c)$. Since η is metric and $(C \cap A) \cup D_n \subset C$, it then follows that

$$\eta(C \cap A) + \eta(D_n) = \eta\big((C \cap A) \cup D_n\big) \leq \eta(C) ,$$

and passing to the limit $n \to \infty$ yields the assertion.

To carry out this train of thoughts we choose a sequence of real numbers $\varepsilon_1 > \varepsilon_2 > \cdots > 0$ converging to zero and set

$$D_n := \{x \in C \cap A^c : |x - y| \geq \varepsilon_n \text{ for all } y \in A\} .$$

Now $a(C \cap A, D_n) \geq \varepsilon_n > 0$ holds, as desired.

In order to prove the other property of the sets D_n, we additionally consider the sets $E_n := D_{n+1} \setminus D_n$, $n \geq 1$. For any $m \geq 1$ we have $a(E_{n+m}, E_{n-1}) \geq \varepsilon_n - \varepsilon_{n+1} > 0$. Since η is metric, it follows that

$$\sum_{k=1}^{n} \eta(E_{2k}) = \eta\Big(\bigcup_{k=1}^{n} E_{2k}\Big) \leq \eta(C)$$

and analogously $\sum_{k=1}^{n} \eta(E_{2k-1}) \leq \eta(C)$, and we obtain $\sum_{k \geq 1} \eta(E_k) < \infty$, because we have assumed $\eta(C) < \infty$ to hold.

Since A is assumed to be closed, we have $C \cap A^c = D_n \cup \bigcup_{m \geq n} E_m$ and consequently, by virtue of σ-subadditivity,

$$\eta(D_n) \leq \eta(C \cap A^c) \leq \eta(D_n) + \sum_{m \geq n} \eta(E_m) .$$

As $n \to \infty$ the rightmost expression tends to 0, and so $\eta(D_n) \to \eta(C \cap A^c)$. Therefore, the sets $D_1, D_2, \ldots$ have the desired properties, and so all closed sets are η-measurable. This holds for all Borel sets, too, because the closed sets generate the Borel σ-algebra. □

The metric outer measures η_s resp. the measures arising from the restriction to the Borel σ-algebra are called *Hausdorff measures*. For geometric investigations one rather uses them as a family, the value of the parameter s is chosen for each $A \subset \mathbb{R}^d$ separately.

Lemma. *For every* $A \subset \mathbb{R}^d$ *there exists a number* $0 \le h_A \le d$ *such that*

$$\eta_s(A) = \begin{cases} \infty\,, & \text{if } s < h_A\,, \\ 0\,, & \text{if } s > h_A\,. \end{cases}$$

Proof. According to the definition of $\eta_{s,\delta}$ we have for all $\varepsilon > 0$

$$\eta_{s+\varepsilon,\delta}(A) \le \delta^\varepsilon \eta_{s,\delta}(A)\,.$$

When $\eta_s(A) < \infty$, passing to the limit $\delta \to 0$ yields $\eta_{s+\varepsilon}(A) = 0$. Thus, a number $h_A \in [0, \infty]$ exists which is related to $\eta_s(A)$ as claimed.

It remains to prove that $h_A \le d$. The unit cube $[0, 1)^d$ can be partitioned in an obvious manner into n^d subcubes of sidelength $1/n$ and diameter $\sqrt{d}/n$. Therefore,

$$\eta_{d,\sqrt{d}/n}\big([0,1)^d\big) \le n^d\big(\sqrt{d}/n\big)^d = d^{d/2}\,,$$

and letting $n \to \infty$ we obtain $\eta_d\big([0,1)^d\big) < \infty$. For every $\varepsilon > 0$ it follows that $\eta_{d+\varepsilon}\big([0,1)^d\big) = 0$, and by virtue of σ-additivity we get $\eta_{d+\varepsilon}(\mathbb{R}^d) = 0$. This proves that $h_A \le d$ for all $A \subset \mathbb{R}^d$. □

The number h_A is called the *Hausdorff dimension* of A. In geometric measure theory the Hausdorff dimension and measure are studied in more detail. It turns out that in all cases for which one can assign a dimension to A in an intuitive manner, this dimension coincides with the Hausdorff dimension. Moreover, in the d-dimensional case the Hausdorff measure for $s = d$ coincides with the Lebesgue measure, except for a positive normalization constant which however is not easy to determine. We do not discuss this further and close this section with an example.

Example (Cantor set)

The Hausdorff dimension of the Cantor set C can be determined easily through a heuristic scaling argument. For any set $A \subset \mathbb{R}$ and any $c > 0$ we set $cA := \{cx : x \in A\}$. Then we have (compare Exercise XI.2)

$$\eta_s(cA) = c^s \eta(A)\,.$$

Obviously $C = C' \cup C''$ for some disjoint sets C' and C'' which result from C by scaling with the factor $c = 1/3$ and by translation. Hence,

$$\eta_s(C) = \eta_s(C') + \eta_s(C'') = 2 \cdot 3^{-s}\eta_s(C) .$$

Assuming that $0 < \eta_h(C) < \infty$ holds for the Hausdorff dimension $h = h_C$ of C, it follows that $1 = 2 \cdot 3^{-h}$ or

$$h = \frac{\log 2}{\log 3} = 0,631 .$$

We now want to show that for this number h we indeed have $1/2 \leq \eta_h(C) \leq 1$. On the one hand, C is contained in C_n, the disjoint union of 2^n intervals of length 3^{-n}. Therefore,

$$\eta_{h,3^{-n}}(C) \leq 2^n(3^{-n})^h = 1$$

and $\eta_h(C) \leq 1$.

For the other estimate we utilize the bijection $\varphi : [0, 1) \to C$ which we have introduced in Chap. 9 in the section dealing with the Cantor set. For all $y, y' \in [0, 1)$ we have

$$2\big|\varphi(y) - \varphi(y')\big|^h \geq |y - y'| .$$

Indeed, let n be the first position in the binary representations $y = \sum_{k\geq 1} y_k 2^{-k}$ and $y' = \sum_{k\geq 1} y'_k 2^{-k}$ where $y_n \neq y'_n$ holds. Then

$$|y - y'| \leq \sum_{k\geq n} 2^{-k} = 2^{-n+1} , \quad |\varphi(y) - \varphi(y')| \geq 2\Big(3^{-n} - \sum_{k>n} 3^{-k}\Big) = 3^{-n} ,$$

and the assertion follows since $(3^{-n})^h = 2^{-n}$. For any interval $A \subset \mathbb{R}$ this yields

$$2d(A)^h \geq d\big(\varphi^{-1}(A)\big) = \lambda\big(\varphi^{-1}(A)\big) .$$

If now $C \subset \bigcup_{m\geq 1} A_m$ holds for some intervals $A_1, A_2, \ldots$, the σ-continuity of the Lebesgue measure and the fact that $[0, 1) \subset \bigcup_{m\geq 1} \varphi^{-1}(A_m)$ imply that

$$2\sum_{m\geq 1} d(A_m)^h \geq \sum_{m\geq 1} \lambda\big(\varphi^{-1}(A_m)\big) \geq 1 .$$

Because in the one-dimensional case it is obviously sufficient to consider coverings consisting of intervals only, we obtain that $\eta_h(C) \geq 1/2$. By the way, a more precise analysis reveals that $\eta_h(C) = 1$.

Exercises

11.1 Let ν be the measure used in the proof of the extension theorem which results from restricting the outer measure η to the σ-algebra $\mathcal{A}_\eta$. Prove that if ν is finite (or, at least, σ-finite), then ν coincides with the completion of μ.
Hint: First show that for each $A \subset S$ there exists an $A' \supset A$ such that $\mu(A') = \eta(A)$. A' can be chosen as having the form $A' = \bigcap_{n\geq 1} \bigcup_{m\geq 1} E_{mn}$ where $E_{mn} \in \mathcal{E}$.

11.2 Prove that the Hausdorff measure satisfies

$$\eta_s(cA) = c^s \eta_s(A) .$$

Conclude that in the d-dimensional case η_s is different from the Lebesgue measure when $s \neq d$, and moreover cannot be made to coincide with it by scaling.

Hilbert Spaces

12

We return to the space $L_2(S;\mu)$ of square integrable functions whose basic properties we have discussed in Chap. 6. They are related to some geometric issues which we now want to learn about. Those are the properties of a Hilbert space,[1] for which the space $L_2(S;\mu)$ serves as a prototype.

A Hilbert space is a vector space where not only a length is associated to each vector, but moreover any two vectors – by means of a scalar product – enclose an angle between them, in particular it is possible to state whether they are orthogonal to each other. Its additional geometric properties make it possible to find, for any convex closed set K and any point x not belonging to K, a point in K whose distance to x is minimal. This leads to widely used orthogonal decompositions, the most prominent being the Fourier series.

We recall the definition of a scalar product in a vector space over the real or complex field. For any $\alpha \in \mathbb{C}$ we denote by $\overline{\alpha}$ the complex number conjugate to α; note that $\alpha\overline{\alpha} = |\alpha|^2$.

Definition

A *scalar product* is a mapping which to any two elements x, y of a vector space X associates a number (x, y) with the following properties:

(i) Positive definiteness: $(x, x) > 0$ for any $x \neq 0$,
(ii) $(y, x) = \overline{(x, y)}$ for all vectors $x, y \in X$,
(iii) $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$ for all $x, y, z \in X$ and all scalars α, β.

[1] David Hilbert, 1862–1943, born at Königsberg, active in Königsberg and Göttingen. The 23 problems presented by him in Paris in 1900 and named after him have deeply influenced the development of mathematics. With him and his activity, which extended to all branches of mathematics, Göttingen became the world center of mathematics.

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks
in Mathematics, DOI 10.1007/978-3-319-15365-0_12

Properties (ii) and (iii) directly imply $(x, 0) = (0, x) = 0$ and

$$(x, \alpha y + \beta z) = \overline{\alpha}(x, y) + \overline{\beta}(x, z)$$

for any vectors x, y, z and any scalars α, β. In the real case, a scalar product thus is just a symmetric positive definite bilinear form.

Example

1. With $(x^1, x^2, \dots)$ and $(y^1, y^2, \dots)$ being two sequences of scalars of length d, the expression

$$(x, y) = \sum_{n=1}^{d} x^n \overline{y^n}$$

   defines a scalar product on the space $\mathbb{R}^d$, resp. $\mathbb{C}^d$, if d is finite; in the case $d = \infty$ we arrive at the space

$$\ell^2 := \Big\{(x^1, x^2, \dots) : \sum_n |x^n|^2 < \infty\Big\}$$

   of all square summable real, resp. complex sequences.
2. The definition

$$(f, g) = \int f\overline{g}\, d\mu$$

   yields a scalar product in the space $L_2(S; \mu)$ of all square integrable functions on a measure space $(S, \mathcal{A}, \mu)$. Here, the integral of a complex-valued function $h = h_1 + ih_2$ with $h_1, h_2 \in L_1(S; \mu)$ is defined as

$$\int h\, d\mu = \int h_1\, d\mu + i \int h_2\, d\mu\,.$$

We set

$$\|x\| := \sqrt{(x, x)}\,.$$

The following fact is known from analysis and linear algebra.

Proposition 12.1. *In a vector space* X *with scalar product* $(\cdot, \cdot)$*,* $\|\cdot\|$ *defines a norm, and there holds the Cauchy-Schwarz inequality*

$$|(x, y)| \le \|x\|\,\|y\| \quad \textit{for all } x, y \in X\,.$$

In particular, $d(x, y) := \|x - y\|$ defines a metric d on X. With respect to this metric we thus may speak of convergence $x_n \to x$ of sequences $x_n \in X$ to a limit $x \in X$, of closed subsets of X, and so on. Since $(x_n, y_n) - (x, y) = (x_n - x, y_n - y) + (x_n - x, y) + (x, y_n - y)$ and therefore $|(x_n, y_n) - (x, y)| \le \|x_n - x\|\|y_n - y\| + \|x_n - x\|\|y\| + \|x\|\|y_n - y\|$, it holds that

$$(x_n, y_n) \to (x, y)\,, \quad \text{if } x_n \to x\,,\ y_n \to y\,.$$

The scalar product is thus continuous, and therefore so is the norm. In the next chapter we will recall the notion of a norm and its implications in more detail.

In a vector space X with scalar product there holds the *parallelogram identity*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)\,,$$

an immediate consequence of the formula $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2 \pm [(x, y) + (y, x)]$. Moreover, the definitions directly imply that in the real case the scalar product satisfies the identity

$$(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2) \qquad (*)$$

for any $x, y \in X$. If, conversely, $\|\cdot\|$ is a norm on X satisfying the parallelogram identity for all $x, y \in X$, then after some computation one sees that, in the real case, $(*)$ indeed defines a scalar product. In the complex case, a different formula is valid (Exercise 12.1). This passage between the square of the norm and the scalar product is called *polarization*.

Two vectors x, y on a Hilbert space X are called *orthogonal* if $(x, y) = 0$. From $(x + y, x + y) = (x, x) + (x, y) + (y, x) + (y, y)$ we obtain for orthogonal vectors $x, y \in X$ the “Pythagoras Theorem”

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2\,.$$

For $M \subset X$, the set

$$M^\perp = \{x : (x, y) = 0 \text{ for all } y \in M\}$$

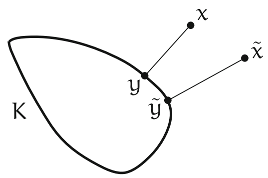
is called the *orthogonal complement* of M. We obviously have $N^\perp \supset M^\perp$ whenever $N \subset M$, as well as $\overline{M}^\perp = M^\perp$, because $(x_n, y) = 0$ for every n and $x_n \to x$ imply that $(x, y) = 0$. We also note that $M^\perp$ is a closed subspace of X.

Definition

A vector space X endowed with a scalar product is called *Hilbert space*, if it is complete w.r.t. the corresponding norm $\|\cdot\|$, that is, if every Cauchy sequence w.r.t. the metric $d(x, y) := \|x - y\|$ is convergent.

The Projection Theorem

If K is a closed convex subset of the plane, for every point x of the plane we can find a unique point y in K which minimizes the distance to x, as the figure illustrates.



This fact remains true in general Hilbert spaces.

Proposition 12.2 (Projection Theorem I). *Let* K *be a closed, convex and nonempty subset of a Hilbert space* X*. Then for every* $x \in X$ *there exists a unique* $y \in K$ *such that*

$$\|x - y\| = \min_{z \in K} \|x - z\| .$$

The point y is called the *projection* of x onto K, written as $y = P_K x$. If $x \in K$ we have $y = x$.

Proof. To prove existence, for any given $x \in X$ we choose a *minimizing sequence* $\{y_n\}$ in K satisfying $\lim_n \|x - y_n\| = \inf_{z \in K} \|x - z\| =: d$. Using the parallelogram identity we get

$$2(\|x - y_n\|^2 + \|x - y_m\|^2) = \|2x - (y_n + y_m)\|^2 + \|y_n - y_m\|^2 .$$

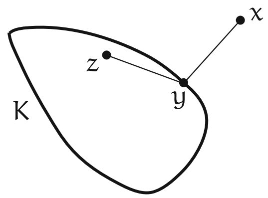
Since $(y_n + y_m)/2 \in K$ by convexity, it follows that $\|x - (y_n + y_m)/2\| \geq d$ and therefore $\|y_n - y_m\|^2 \leq 2(\|x - y_n\|^2 + \|x - y_m\|^2) - 4d^2 \to 0$ as $n, m \to \infty$. Thus, y_n is a Cauchy sequence. Since X is complete, there exists $y = \lim_n y_n$. As K is closed, y belongs to K, and the continuity of the norm implies that $\|x - y\| = \lim_n \|x - y_n\| = d$.

To prove uniqueness, let $\tilde{y} \in K$ with $\|x - \tilde{y}\| = d$. As above, the parallelogram identity implies that

$$\|y - \tilde{y}\|^2 = 2(\|x - y\|^2 + \|x - \tilde{y}\|^2) - \|2x - y - \tilde{y}\|^2 = 4d^2 - 4\|x - (y + \tilde{y})/2\|^2 \leq 0 ,$$

because $(y + \tilde{y})/2 \in K$. This yields $y = \tilde{y}$. □

As the next figure shows, the angle between the difference vectors $x - y$ and $z - y$ amounts to at least 90°, for any $z \in K$.



The projection can be characterized by this property.

Proposition 12.3 (Projection Theorem II). *Let* K *be a closed, convex and nonempty subset of a real Hilbert space* X*. Then for every* $x \in X$ *there is a unique solution* $y \in K$ *of the inequalities*

$$(x - y, z - y) \leq 0 \quad \textit{for all } z \in K\,, \tag{$*$}$$

and it holds $y = P_K x$.

The system $(*)$ of inequalities is usually called a *variational inequality*. One may interpret it as the variational form of an inequality for the vector $x - y$.

Proof. If $y, \tilde{y} \in K$ solve $(*)$ for a given $x \in X$, we have $(x - y, \tilde{y} - y) \leq 0$ as well as $(x - \tilde{y}, y - \tilde{y}) \leq 0$. Adding these inequalities yields $0 \geq (x - y - x + \tilde{y}, \tilde{y} - y) = -\|\tilde{y} - y\|^2$, which implies uniqueness. We prove that $y = P_K x$ is a solution. For any $z \in K$ and any $t \in (0, 1)$ we have $z_t := (1 - t)y + tz \in K$, thus we obtain due to $x - z_t = (x - y) + t(y - z)$ that

$$\|x - y\|^2 \leq \|x - z_t\|^2 = \|x - y\|^2 + 2(x - y, t(y - z)) + t^2\|z - y\|^2$$

and therefore $0 \leq 2(x - y, y - z) + t\|z - y\|^2$ after dividing by t. Passing to the limit $t \to 0$ yields the assertion. □

In the case of a complex Hilbert space, too, the projection is characterized by a variational inequality, namely

$$\operatorname{Re}(x - y, z - y) \leq 0 \quad \text{for all } z \in K\,.$$

The proof is analogous.

Let $y = P_K x$ and $\tilde{y} = P_K \tilde{x}$ be the projections of the two points $x, \tilde{x} \in X$. Adding the inequalities $\mathrm{Re}\,(x - y, \tilde{y} - y) \le 0$ and $\mathrm{Re}\,(\tilde{x} - \tilde{y}, y - \tilde{y}) \le 0$ and using the Cauchy-Schwarz inequality we get

$$\|\tilde{y} - y\|^2 = (\tilde{y} - y, \tilde{y} - y) \le \mathrm{Re}\,(\tilde{x} - x, \tilde{y} - y) \le \|\tilde{x} - x\| \, \|\tilde{y} - y\| \, ,$$

and therefore

$$\|P_K \tilde{x} - P_K x\| \le \|\tilde{x} - x\| \, .$$

This means that the projection $P_K : X \to K$ is Lipschitz continuous. Since $P_K x = x$ for $x \in K$, the Lipschitz constant equals 1 if K consists of more than a single point. One says that the projection P_K is non-expansive.

If in particular $K = U$ is a closed subspace of X, the variational inequality turns into the *variational equation*

$$(x - y, v) = 0 \quad \text{for all } v \in U \, .$$

We obtain it by inserting $z = y \pm v$ and, in the complex case, also $z = y \pm iv$ into the variational inequality. The projection P_U is linear in that case, because the variational inequalities for $y = P_U x$ and $\tilde{y} = P_U \tilde{x}$ and for arbitrary scalars α and β immediately imply the variational equation

$$([\alpha x + \beta \tilde{x}] - [\alpha y + \beta \tilde{y}], v) = 0 \quad \text{for all } v \in U \, ,$$

and therefore $P_U(\alpha x + \beta \tilde{x}) = \alpha P_U x + \beta P_U \tilde{x}$. Summarizing these considerations we obtain the following result.

Lemma. *Let* U *be a closed subspace of a Hilbert space* X. *Then the projection* P_U *is a linear continuous mapping from* X *to* U.

Example

If U is a closed subspace of the real Hilbert space $L_2(S; \mu)$, and if $f \in L_2(S; \mu)$, then by the projection theorem $P_U f$ equals the uniquely determined function in U which satisfies

$$\int fg \, d\mu = \int P_U f \cdot g \, d\mu \quad \text{for all } g \in U \, . \qquad (*)$$

A certain special case is of relevance in probability theory. Let μ be a probability measure on $(S, \mathcal{A})$, let $\mathcal{A}'$ be a σ-algebra with $\mathcal{A}' \subset \mathcal{A}$. For $L_2(S; \mu) =: L_2(S; \mathcal{A}, \mu)$ we consider the subspace $U = L_2(S; \mathcal{A}', \mu)$ of all real-valued functions which are square integrable on S and measurable w.r.t. $\mathcal{A}'$. Since U itself is a Hilbert space, U is closed in $L_2(S;\mu)$. The characterization (∗) of the projection given in this example can be written equivalently (Exercise 12.2) as

$$\int_{A'} f\,d\mu = \int_{A'} P_U f\,d\mu \quad \text{for all } A' \in \mathcal{A}' . \tag{$**$}$$

Together with the $\mathcal{A}'$-measurability of $P_U f$, (∗∗) thus asserts that $P_U f$ is the *conditional expectation* of f.

We return to the general situation of a closed subspace U in a Hilbert space X. The variational equation

$$(x - P_U x, v) = 0 \quad \text{for all } v \in U$$

means that $x - P_U x$ is orthogonal to U, thus $x - P_U x \in U^\perp$. Therefore, P_U also is called *orthogonal projection*. Setting $u = P_U x$ and $u^\perp = x - u$, we thus obtain an orthogonal decomposition

$$x = u + u^\perp , \quad u \in U , \quad u^\perp \in U^\perp , \tag{$*$}$$

for which by Pythagoras

$$\|x\|^2 = \|u\|^2 + \|u^\perp\|^2 .$$

Proposition 12.4 (Orthogonal Decomposition). *Let* U *be a closed subspace of a Hilbert space* X*. Each* $x \in X$ *can be decomposed uniquely in the form* (∗)*, and one has* $u = P_U x$ *and* $u^\perp = P_{U^\perp} x$*, as well as*

$$\|P_U x\| \le \|x\| .$$

Proof. We have $U \cap U^\perp = \{0\}$ since $(v, v) = 0$ and therefore $v = 0$ holds for all $v \in U \cap U^\perp$. This implies uniqueness, because for any two such decompositions $x = u + u^\perp = \tilde{u} + \tilde{u}^\perp$ we get $u - \tilde{u} = \tilde{u}^\perp - u^\perp \in U \cap U^\perp$. It remains to show that $u^\perp = P_{U^\perp} x$. For any $w \in U^\perp$ we have $(x - u^\perp, w) = (u, w) = 0$, therefore $u^\perp$ solves the variational equation which characterizes $P_{U^\perp} x$. □

The orthogonal projection enables us to characterize the continuous linear functionals on a Hilbert space. By a continuous linear functional we mean a continuous and linear mapping ℓ from X to the scalar field $\mathbb{R}$ or $\mathbb{C}$. The set of all those functionals ℓ

forms the dual space X' of X, we will treat it in more detail in the next chapter. For a given $\ell \in X'$ one defines

$$\|\ell\| := \sup_{\|x\|\le 1} |\ell(x)| .$$

There holds the following characterization which goes back to F. Riesz.

Proposition 12.5 (Riesz Representation Theorem). *Let* y *be an element of a Hilbert space* X. *Then*

$$x \mapsto (x, y)$$

defines a continuous linear functional. Conversely, any $\ell \in X'$ *can be represented in the form* $\ell(x) = (x, y)$ *for some* $y \in X$. *Here,* y *is uniquely determined by* ℓ, *and* $\|\ell\| = \|y\|$.

Proof. We already know that the first part of the assertion is true. For the converse we consider for any given $\ell \in X'$ its kernel $U = \ell^{-1}(\{0\})$ which is a closed subspace of X since ℓ is continuous. If $\ell = 0$ then we must have $y = 0$, otherwise we may choose some $w \in U^{\perp}$ with $\ell(w) = 1$. For every $x \in X$ we have $x - \ell(x)w \in U$, since $\ell(x - \ell(x)w) = \ell(x) - \ell(x)\ell(w) = 0$. We then get

$$(x, w) = (x - \ell(x)w, w) + (\ell(x)w, w) = \ell(x)\|w\|^2 .$$

The vector $y = \|w\|^{-2}w$ thus has the required property. If on the other hand $0 = (x, y) - (x, \tilde{y}) = (x, y - \tilde{y})$ for all $x \in X$, then in particular $0 = (y - \tilde{y}, y - \tilde{y})$ and therefore $y = \tilde{y}$. Finally we get $\|\ell\| = \|y\|$, since $|\ell(x)| \le \|y\|\|x\| \le \|y\|$ whenever $\|x\| \le 1$, and $\ell(y/\|y\|) = \|y\|$ if $y \ne 0$. □

With the preceding result we also have characterized all closed hyperplanes H in a Hilbert space, because such hyperplanes coincide with the level sets $\{\ell = c\}$ associated to continuous linear functionals.

Corollary. *Let* μ *be a finite measure on a measurable space* $(S, \mathcal{A})$. *Every continuous linear functional* ℓ *on the real Hilbert space* $L_2(S; \mu)$ *is of the form*

$$\ell(f) = \int fg \, d\mu$$

for some $g \in L_2(S; \mu)$, *and* $\|\ell\| = \|g\|_2$.

In the following chapter we will generalize this result to the spaces $L_p(S;\mu)$ for $1 \le p < \infty$.

Bases in Hilbert Spaces

When X is a vector space, a (vector space) basis B is a system of linearly independent vectors in X, such that every $x \in X$ can be represented uniquely as a linear combination

$$x = \sum_{b\in B} \alpha_b b \qquad (*)$$

with finitely many nonzero scalars α_b. This notion, while being central for the treatment of finite-dimensional spaces, is largely useless in the context of infinite-dimensional spaces. Instead, one considers representations for which (*) becomes a series that converges in a suitable sense. The situation is particularly neat in Hilbert space, because one has the scalar product at one's disposal, and thus one is able to form orthonormal systems.

Definition

A subset E of a Hilbert space X is called *orthonormal system*, if $\|e\| = 1$ for each $e \in E$ and $(e, f) = 0$ for all $e, f \in E$ with $e \ne f$.

Example

1. In the space ℓ^2 of square-summable sequences, the set $E = \{e_k : k \in \mathbb{N}\}$ of all unit vectors ($e_k^j = \delta_{kj}$) is an orthonormal system.
2. We consider the space $L_2(-\pi,\pi) := L_2((-\pi,\pi);\lambda)$ and write more precisely $L_2^{\mathbb{C}}(-\pi,\pi)$ and $L_2^{\mathbb{R}}(-\pi,\pi)$ in order to specify the field of scalars. The set $E = \{e_k : k \in \mathbb{Z}\}$, where

$$e_k(t) = \frac{1}{\sqrt{2\pi}} e^{ikt} ,$$

is an orthonormal system in $L_2^{\mathbb{C}}(-\pi,\pi)$, since for any $k \ne j$

$$(e_k, e_j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(k-j)t}\,dt = \frac{1}{2\pi}\frac{1}{i(k-j)} e^{i(k-j)t}\Big|_{t=-\pi}^{t=\pi} = 0$$

and obviously $(e_k, e_k) = 1$. The set $E = \{\tilde{e}_k : k \in \mathbb{Z}\}$, where

$$\tilde{e}_0(t) = \frac{1}{\sqrt{2\pi}} , \quad \tilde{e}_k(t) = \frac{1}{\sqrt{\pi}}\cos kt , \quad \tilde{e}_{-k}(t) = \frac{1}{\sqrt{\pi}}\sin kt , \quad k \ge 1 ,$$

is also an orthonormal system in $L_2^{\mathbb{C}}(-\pi, \pi)$ and therefore, too, in $L_2^{\mathbb{R}}(-\pi, \pi)$, as one sees from the formulas $\tilde{e}_0 = e_0$,

$$\tilde{e}_k = \frac{1}{\sqrt{2}}(e_k + e_{-k}) , \quad \tilde{e}_{-k} = \frac{1}{i\sqrt{2}}(e_k - e_{-k}) , \quad k \geq 1 ,$$

and from the properties of the scalar product, or directly via partial integration.

If $\alpha_1, \ldots, \alpha_n$ are scalars and $e_1, \ldots, e_n$ are distinct elements of an orthonormal system E, then

$$\Big\| \sum_{k=1}^{n} \alpha_k e_k \Big\|^2 = \sum_{k=1}^{n} |\alpha_k|^2 .$$

Indeed, $(\sum_{k=1}^n \alpha_k e_k, \sum_{l=1}^n \alpha_l e_l) = \sum_{k=1}^n \sum_{l=1}^n \alpha_k \overline{\alpha}_l (e_k, e_l) = \sum_{k=1}^n \alpha_k \overline{\alpha}_k$. If $\sum_{k=1}^n \alpha_k e_k = 0$, we must have $\alpha_1 = \cdots = \alpha_n = 0$. Therefore, any orthonormal system is linearly independent.

In addition, orthonormal systems are useful for forming convergent series. A series $\sum_{k \geq 1} y_k$ in a Hilbert space (or, more general, in a normed space) X is called *convergent*, if the sequence $s_n = \sum_{k=1}^n y_k$ of the partial sums converges in X. The limit $y = \lim_n s_n$ is also denoted by $\sum_{k \geq 1} y_k$.

Lemma. *Let $\{e_1, e_2, \ldots\}$ be a countably infinite orthonormal system in the Hilbert space X, and let $\alpha_1, \alpha_2, \ldots$ be a scalar sequence. Then $\sum_{k \geq 1} \alpha_k e_k$ converges in X if and only if $\sum_{k \geq 1} |\alpha_k|^2 < \infty$. We then have*

$$\Big\| \sum_{k \geq 1} \alpha_k e_k \Big\|^2 = \sum_{k \geq 1} |\alpha_k|^2 .$$

Proof. For $s_n := \sum_{k=1}^n \alpha_k e_k$ and any $m < n$ we have

$$\|s_n - s_m\|^2 = \Big\| \sum_{k=m+1}^{n} \alpha_k e_k \Big\|^2 = \sum_{k=m+1}^{n} |\alpha_k|^2 .$$

Therefore, $\{s_n\}$ is a Cauchy sequence if and only if $\sum_{k \geq 1} |\alpha_k|^2$ converges. As X is complete, this is equivalent to the first assertion.

The second assertion results from the fact that $\|s_n\|^2 = \sum_{k=1}^n |\alpha_k|^2$, passing to the limit $n \to \infty$ in view of the continuity of the norm. □

To apply the lemma we utilize the following proposition.

Proposition 12.6 (Bessel's[2] Inequality). *Let* $\{e_1, e_2, \dots\}$ *be a finite or countably infinite orthonormal system in* X. *Then for every* $x \in X$ *we have*

$$\sum_{k\geq 1} |(x, e_k)|^2 \leq \|x\|^2 .$$

Proof. Setting $s_n := \sum_{k=1}^n (x, e_k)e_k$ we get $\|s_n\|^2 = \sum_{k=1}^n |(x, e_k)|^2$, and therefore

$$(x, s_n) = \sum_{k=1}^{n} \overline{(x, e_k)}(x, e_k) = \|s_n\|^2 = (s_n, s_n) .$$

It follows that $(x - s_n, s_n) = 0$, and the Pythagoras Theorem yields

$$\|x\|^2 = \|s_n\|^2 + \|x - s_n\|^2 \geq \|s_n\|^2 = \sum_{k=1}^{n} |(x, e_k)|^2 .$$

This proves the assertion for the finite case and, letting $n \to \infty$, also for the infinite case. □

The Bessel inequality in combination with the foregoing lemma implies that the expression $\sum_{k\geq 1}(x, e_k)e_k$ has a well-defined value for any $x \in X$ and any orthonormal system $\{e_1, e_2, \dots\}$, either as a finite sum or as a convergent series. We may interpret those expressions as projections on subspaces.

For this purpose, let span(E) denote the subspace of X spanned by the set $E \subset X$. It consists of all linear combinations of the form $\sum_{e\in E} \alpha_e e$ with scalars α_e, only finitely many of them being nonzero. We denote its closure by $\overline{\text{span}}(E)$. If E is finite, then $\text{span}(E) = \overline{\text{span}}(E)$, because every finite-dimensional normed space is complete. If E is infinite, a result of functional analysis says that $\text{span}(E) \neq \overline{\text{span}}(E)$ if X is complete. (If that would not be true, one might work with E as a vector space basis.)

Proposition 12.7. *Let* U *be a closed subspace and* $E = \{e_1, e_2, \dots\}$ *a finite or countably infinite orthonormal system in the Hilbert space* X, *assume that* $\overline{\text{span}}(E) = U$. *Then for all* $x \in X$ *one has that*

$$P_U x = \sum_{k\geq 1} (x, e_k)e_k , \quad \|P_U x\|^2 = \sum_{k\geq 1} |(x, e_k)|^2 .$$

[2] Friedrich Wilhelm Bessel, 1784–1846, born in Minden, active at the observatory in Königsberg. He worked in astronomy, mathematics, and geodesics.

Proof. Let $y := \sum_{k \geq 1}(x, e_k)e_k$. It then holds that (in the finite case and also, due to the continuity of the scalar case, in the infinite case)

$$(x - y, e_l) = (x, e_l) - \sum_{k \geq 1}(x, e_k)(e_k, e_l) = 0 \, .$$

It follows that $(x - y, z) = 0$ for all $z \in \operatorname{span}(E)$ and, due to the continuity of the scalar product, $(x - y, z) = 0$ for all $z \in \overline{\operatorname{span}}(E) = U$. Therefore, y satisfies the variational equation characterizing the projection, and the first assertion follows. The second assertion is a consequence of the foregoing lemma. □

Example

1. In the sequence space ℓ^2 we consider the orthonormal system $E = \{e_k : k \in \mathbb{N}\}$ consisting of the standard unit vectors. For any $x = (x^1, x^2, \dots) \in \ell^2$ we have $(x, e_k) = x^k$, and

$$P_U x = \sum_{k=1}^{n}(x, e_k)e_k = \sum_{k=1}^{n} x^k e_k$$

   is the orthogonal projection onto $U = \operatorname{span}(\{e_1, \dots, e_n\})$.
2. In the function space $L_2^{\mathbb{C}}(-\pi, \pi)$ we investigate the orthonormal system given by the functions $e_k(t) = (1/\sqrt{2\pi})e^{ikt}$. For any $f \in L_2^{\mathbb{C}}(-\pi, \pi)$,

$$c_k = (f, e_k) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)e^{-ikt} \, dt \, , \quad k \in \mathbb{Z} \, ,$$

   is called the k-*th Fourier coefficient*[3] *of* f. Setting $U = \operatorname{span}(\{e_{-n}, \dots, e_n\})$, the orthogonal projection

$$P_U f = \sum_{k=-n}^{n} (f, e_k)e_k = \sum_{k=-n}^{n} c_k e_k$$

   is just the n-th partial sum of the *Fourier series* $\sum_{k \in \mathbb{Z}} c_k e_k$ of f. Concerning the convergence of the Fourier series, the results below will provide information.

We now arrive at the notion which in Hilbert space replaces the notion of a vector space basis.

[3] JOSEPH FOURIER, 1768–1830, born in Auxerre, active in Paris at the École Polytechnique. In the context of his fundamental contribution to heat conduction he utilized, for the first time, trigonometric series for the representation of general functions.

Definition

An orthonormal system E is called *orthonormal basis*[4] of X, if span(E) is dense in X, that is, if $\overline{\text{span}}(E) = X$.

Thus one only requires that every $x \in X$ can be represented as the limit of a sequence in span(E). Consequently, in Hilbert space any element x can be represented as the limit of a series whose partial sums belong to span(E).

Proposition 12.8. *For a countably infinite orthonormal system* $E = \{e_1, e_2 \dots\}$ *in a Hilbert space* X *the following assertions are equivalent:*

(i) $E^\perp = \{0\}$.

(ii) $X = \overline{\text{span}}(E)$*, that is,* E *is an orthonormal basis.*

(iii) *There holds*

$$x = \sum_{k=1}^{\infty} (x, e_k) e_k \quad \textit{for all } x \in X\,.$$

(iv) *There holds*

$$(x, y) = \sum_{k=1}^{\infty} (x, e_k)(e_k, y) \quad \textit{for all } x, y \in X\,.$$

(v) *The* Parseval[5] identity holds:

$$\|x\|^2 = \sum_{k=1}^{\infty} |(x, e_k)|^2 \quad \textit{for all } x \in X\,.$$

Proof. (i) ⇒ (ii): Let $U := \overline{\text{span}}(E)$. From $E \subset U$ it follows that $U^\perp \subset E^\perp$, thus $U^\perp = \{0\}$ and therefore $U = X$.

(ii) ⇒ (iii): From $U := \overline{\text{span}}(E) = X$ it follows that $P_U x = x$ for all $x \in X$ and thus the assertion, by the preceding proposition.

(iii) ⇒ (iv): The series on the right side of (iv) converges absolutely, since

$$\sum_{k=1}^{\infty} |(x, e_k)(e_k, y)| \le \sum_{k=1}^{\infty} |(x, e_k)|^2 \cdot \sum_{k=1}^{\infty} |(y, e_k)|^2 \le \|x\|^2 \|y\|^2$$

[4]Instead of an orthonormal basis one also speaks of a *complete orthonormal system*.

[5]Marc-Antoine Parseval, 1755–1836, born in Rosière-aux-Salines, active in Paris.

due to the Cauchy-Schwarz inequality in ℓ^2 and the Bessel inequality. The assertion now follows when on both sides of (iii) we take the scalar product with y and use the continuity of the scalar product.

(iv) $\Rightarrow$ (v): We set $y = x$ in (iv).

(v) $\Rightarrow$ (i): For every $x \in E^{\perp}$ we have $(x, e_k) = 0$ for all k, and therefore $\|x\| = 0$ by (v). □

Example

In the sequence space ℓ^2, the orthonormal system $E = \{e_k : k \in \mathbb{N}\}$ consisting of the standard unit vectors is an orthonormal basis, since for any $x = (x^1, x^2, \dots) \in \ell^2$ the sequence given by $s_n = \sum_{k \le n} x^k e_k$ belongs to span (E) and converges to x; thus condition (ii) in the preceding proposition is satisfied.

In order to prove that the orthonormal system given by the functions $e_k(t) = (1/\sqrt{2\pi})e^{ikt}$ actually is an orthonormal basis of $L_2^{\mathbb{C}}(-\pi, \pi)$, we utilize arguments from analysis. We owe to Fejér the idea of investigating, instead of the sequence of partial sums $s_n = \sum_{|k| \le n}(f, e_k)e_k$, the sequence defined by their arithmetic means

$$a_m := \frac{1}{m+1} \sum_{n=0}^{m} \sum_{k=-n}^{n} (f, e_k)e_k \,.$$

Proposition 12.9 (Fejér[6]). *Let* $f : [-\pi, \pi] \to \mathbb{C}$ *be a continuous function satisfying* $f(-\pi) = f(\pi)$. *Then* a_m *converges uniformly to* f *on* $[-\pi, \pi]$.

Proof. We have

$$a_m(t) = \frac{1}{m+1} \sum_{n=0}^{m} \sum_{k=-n}^{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\tau)e^{-ik\tau}\, d\tau \cdot e^{ikt} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\tau)F_m(t-\tau)\, d\tau \,,$$

with the *Fejér kernel*

$$F_m(\tau) = \frac{1}{m+1} \sum_{n=0}^{m} \sum_{k=-n}^{n} e^{ik\tau} \,.$$

[6] Lipót Fejér, 1880–1959, born in Pécs, active in Klausenburg and Budapest. He worked in harmonic analysis and potential theory.

We have $\int_{-\pi}^{\pi} F_m(\tau)\,d\tau = 2\pi$, since $\int_{-\pi}^{\pi} e^{ik\tau}\,d\tau = 0$ for any $k \neq 0$. Since F_m is a 2π-periodic function,

$$a_m(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t-\tau)F_m(\tau)\,d\tau\,,$$

where we have extended f periodically outside $[-\pi, \pi]$. As $f(\pi) = f(-\pi)$, this extension preserves continuity.

Due to a certain trigonometric identity (Exercise 12.4) we have

$$F_m(\tau) = \frac{1}{m+1}\frac{\sin^2(\frac{m+1}{2}\tau)}{\sin^2(\frac{1}{2}\tau)}\,. \qquad (*)$$

For any $0 < \delta < \pi$ we estimate on $[-\pi, \pi]$:

$$|f(t) - a_m(t)| = \frac{1}{2\pi}\Big|\int_{-\pi}^{\pi}(f(t) - f(t-\tau))F_m(\tau)\,d\tau\Big|$$
$$\leq \frac{1}{2\pi}\int_{-\delta}^{\delta}|f(t) - f(t-\tau)|F_m(\tau)\,d\tau + \frac{1}{2\pi}\Big(\int_{-\pi}^{-\delta} + \int_{\delta}^{\pi}\Big)|f(t) - f(t-\tau)|F_m(\tau)\,d\tau\,.$$

For any given $\varepsilon > 0$ we choose $\delta > 0$, according to the uniform continuity of f, such that $|f(t) - f(t-\tau)| < \varepsilon$ for any $|\tau| < \delta$, and we choose m_0 by virtue of $(*)$ such that $F_m(\tau) < \varepsilon$ for all τ satisfying $\delta \leq |\tau| \leq \pi$ and all $m \geq m_0$. It follows that

$$\|f - a_m\|_\infty \leq (1 + 2\|f\|_\infty)\varepsilon$$

for all $m \geq m_0$, and thus the assertion is proved. □

Corollary. *The functions* $e_k(t) = (1/\sqrt{2\pi})e^{ikt}$, $k \in \mathbb{Z}$, *constitute an orthonormal basis of* $L_2^{\mathbb{C}}(-\pi, \pi)$. *The functions*

$$\tilde{e}_0(t) = \frac{1}{\sqrt{2\pi}}\,,\quad \tilde{e}_k(t) = \frac{1}{\sqrt{\pi}}\cos kt\,,\quad \tilde{e}_{-k}(t) = \frac{1}{\sqrt{\pi}}\sin kt\,,\quad k \geq 1\,,$$

constitute an orthonormal basis of $L_2^{\mathbb{R}}(-\pi, \pi)$.

Proof. Let $U = \operatorname{span}\{e_k : k \in \mathbb{Z}\}$. Then $a_m \in U$ and $\|f - a_m\|_2 \leq \sqrt{2\pi}\|f - a_m\|_\infty$. By Fejér's result, U is dense in the subspace V of $L_2^{\mathbb{C}}(-\pi, \pi)$ consisting of the continuous functions satisfying $f(\pi) = f(-\pi)$. Modifying it near a boundary point, we may approximate any arbitrary continuous function by functions from V with an arbitrarily small error in the L_2 norm, and because the continuous functions are dense in $L_2^{\mathbb{C}}(-\pi, \pi)$ by Proposition 7.7, this remains valid for arbitrary functions from L_2. Thus, condition (ii) in the proposition characterizing orthonormal bases holds. Since the Fejér kernel F_m is real-valued, und thus a_m, too, is real-valued for any real-valued f, and because every real-valued function in U is a real linear combination of the functions $\tilde{e}_k$, the assertion also follows for $L_2^{\mathbb{R}}(-\pi, \pi)$. □

Summarizing, for functions $f \in L_2$ the Fourier series $\sum_{k\in\mathbb{Z}}(f, e_k)e_k$ converges to f in the sense of the norm of L_2.

It is a famous result due to Carleson that for any $f \in L_2$ the whole Fourier series converges to f almost everywhere (and not only a subsequence of the partial sums, a consequence of Propositions 6.4 and 6.6).

Exercises

12.1 Prove that in a complex Hilbert space X, the scalar product satisfies the identity

$$(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2)$$

for all $x, y \in X$.

12.2 Let μ be a probability measure on $(S, \mathcal{A})$, let $\mathcal{A}' \subset \mathcal{A}$ be another σ-algebra, let $X = L_2(S; \mathcal{A}, \mu)$ and $U = L_2(S; \mathcal{A}', \mu)$. Prove that for every $f \in X$ and every $h \in U$,

$$\int_{A'} f\, d\mu = \int_{A'} P_U f\, d\mu \quad \text{for all } A' \in \mathcal{A}'$$

implies that $h = P_U f$.

12.3 Let $E = \{e_1, e_2, \ldots\}$ be a countably infinite orthonormal system in a Hilbert space X, let $x \in X$. Then

$$\sum_{k\geq 1}(x, e_{\pi(k)})e_{\pi(k)} = \sum_{k\geq 1}(x, e_k)e_k$$

for any reordering of E given by a bijective mapping $\pi : \mathbb{N} \to \mathbb{N}$. This property is termed *unconditional convergence* of the series $\sum_{k\geq 1}(x, e_k)e_k$.

12.4 Prove that for every $-\pi \leq \tau \leq \pi$, $\tau \neq 0$ the trigonometric identity

$$\sum_{n=0}^{m}\sum_{k=-n}^{n} e^{ikt} = \frac{\sin^2(\frac{m+1}{2}\tau)}{\sin^2(\frac{1}{2}\tau)}$$

is valid. Hint: Use the formula giving the partial sums of the geometric series as well as the trigonometric identity

$$4\sin^2\varphi = 4\Big[\frac{1}{2i}(e^{i\varphi} - e^{-i\varphi})\Big]^2 = 2 - e^{2i\varphi} - e^{-2i\varphi}\,.$$

Banach Spaces

13

In the preceding chapters, several times already we have interpreted functions as elements of function spaces. We now deepen this view, looking more closely at continuous linear functionals on such spaces. We will characterize them in two important cases intimately linked to integration theory, namely for the spaces of p-integrable functions and of continuous functions. The notion of a Banach spaces provides the appropriate functional analytic framework.

To this end we first want to acquaint the reader a little closer with Banach spaces. We recall the definition of a norm in a vector space.

Definition

A *norm* is a mapping on a real or complex vector space X which to each vector $x \in X$ associates a nonnegative number $\|x\|$ with the properties

(i) Definiteness: $\|x\| = 0$ if and only if $x = 0$.
(ii) Positive homogeneity: $\|\alpha x\| = |\alpha| \|x\|$ for all scalars α.
(iii) Triangle inequality: $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in X$.

We call X, or more precisely $(X, \|\cdot\|)$, a *normed space*. If (ii) and (iii) hold, but not necessarily (i), we speak of a *seminorm* on X.

From the triangle inequality, due to $\|x\| \le \|x - y\| + \|y\|$ we immediately obtain the *reverse triangle inequality*

$$\big|\|x\| - \|y\|\big| \le \|x - y\| , \quad x, y \in X .$$

Example

As shown in Chap. 6, the spaces $L_p(S; \mu)$ of the p-integrable ($1 \le p < \infty$) resp., in the case $p = \infty$, measurable and essentially bounded (equivalence classes of) functions on a measure

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0_13

space $(S, \mathcal{A}, \mu)$ are real normed spaces equipped with the *p-norms*

$$\|f\|_p = \Big(\int |f|^p \, d\mu\Big)^{1/p}, \quad 1 \le p < \infty, \quad \|f\|_\infty = N_\infty(f).$$

If we admit complex-valued functions f, we likewise obtain complex normed spaces. (A complex-valued function is called measurable, resp. integrable if its real and its imaginary part are measurable, resp. integrable. The properties of the norm are proved in the same way as in the real case.)

The spaces $\mathbb{R}^d$ and $\mathbb{C}^d$ with $d < \infty$, equipped with the norms

$$\|x\|_p = \Big(\sum_{k=1}^{d} |x^k|^p\Big)^{1/p}, \quad 1 \le p < \infty, \quad \|x\|_\infty = \sup_k |x^k|,$$

can be viewed as special cases of L_p spaces. For μ we choose the counting measure on $S = \{1, \dots, d\}$; here x^k denotes the k-th component of the vector x. In the case $d = 1$ we obtain the scalar field, interpreted as a normed space with $\|x\| = |x|$.

For $d = \infty$ we obtain the spaces ℓ^p of sequences which are summable to the p-th power resp. bounded, consisting of those sequences $x = (x^1, x^2, \dots)$ for which $\|x\|_p$ is finite. With the choice $S = \mathbb{N}$ and the counting measure for μ, they too become special cases of the spaces $L^p(\mu)$.

Example

For an arbitrary set S, the vector space of all bounded (real- or complex-valued) functions on S is a normed space if equipped with $\|f\|_\infty = \sup_{x \in S} |f(x)|$. When S is a compact metric space, the same definition yields a norm on the vector space C(S) of all continuous functions on S, too.

The preceding example illustrates the fact that every subspace U of a normed space X becomes a normed space if we restrict the norm on X to U.

Definition

A sequence $x_1, x_2, \dots$ in a normed space X is said to *converge to the limit* $x \in X$, written as

$$x = \lim_{n\to\infty} x_n, \quad \text{or } x_n \to x,$$

if $\lim_{n\to\infty} \|x_n - x\| = 0$.

In the case $X = L_p(S; \mu)$ for $1 \le p < \infty$ this is just the convergence in p-mean. Convergence in the sup-norm $\|\cdot\|_\infty$ in a function space is synonymous to uniform convergence (resp. to uniform convergence almost everywhere).

It is an immediate consequence of the definition of a norm that sums and scalar multiples of convergent sequences converge to the sum, resp. multiple of their limits.

In a normed space X, the closed ball of radius $r > 0$ around a point $x \in X$ is given by $\{y : \|y - x\| \le r\}$, we denote it by $B_r(x)$. Instead of $B_r(0)$ we briefly write B_r, instead of B_1 simply B; the latter is called the (closed) *unit ball* in X. We obviously have $B_r(x) = x + rB$.

Setting $d(x, y) = \|x - y\|$, any norm on a vector space X generates a translation invariant metric, that is, it holds that $d(x + z, y + z) = d(x, y)$ for all $x, y, z \in X$. Via restriction of d, any subset M of X becomes a metric space.

Definition

A complete normed space is called a *Banach space*.

The above-mentioned spaces $L_p(S; \mu)$ (S measure space) and C(S) (S compact metric space) are Banach spaces. For L_p we have proved this in Chap. 6. To prove the completeness of C(S) one shows that every Cauchy sequence of continuous functions converges uniformly to its pointwise limit (Exercise 13.1).

Setting $\|x\| := \sqrt{(x, x)}$, every Hilbert space becomes a Banach space. A subspace U of a Banach space X obviously is itself a Banach space if and only if it is closed in X. Every finite-dimensional normed space (and therefore every finite-dimensional subspace of a normed space, too) is a Banach space (Exercise 13.3).

The reverse triangle inequality $|\|x\| - \|y\|| \le \|x - y\|$ says that the norm is a Lipschitz continuous function on X with Lipschitz constant equal to 1. Such functions are called *non-expansive*. The *distance function*, defined for $M \subset X$ and $x \in X$ by

$$d(x, M) = \inf_{z \in M} d(x, z) = \inf_{z \in M} \|x - z\| ,$$

when viewed as a function of x, too is non-expansive (Exercise 13.2).

Definition

Two norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on a vector space X are called *equivalent* if there exist constants $c_1, c_2 > 0$ such that

$$c_1 \|x\|_a \le \|x\|_b \le c_2 \|x\|_a , \quad \text{for all } x \in X.$$

This notion indeed yields an equivalence relation on the set of all norms on X, as one immediately checks. Whenever two norms are equivalent, they generate the same topology, that is, in either norm the same sequences are convergent, the same sets are open and closed, and so on.

Proposition 13.1. *On* $\mathbb{R}^d$ *and* $\mathbb{C}^d$, $d \in \mathbb{N}$, *all norms are equivalent.*

Proof. It suffices to show that each norm $\|\cdot\|$ is equivalent to the supremum norm $\|\cdot\|_\infty$. Let e_i be the standard unit vectors in X, $X = \mathbb{R}^d$ or $\mathbb{C}^d$. Then

$$\|x\| \leq \sum_{i=1}^{d} |x_i| \|e_i\| \leq c_2 \|x\|_\infty , \quad c_2 := \sum_{i=1}^{d} \|e_i\| .$$

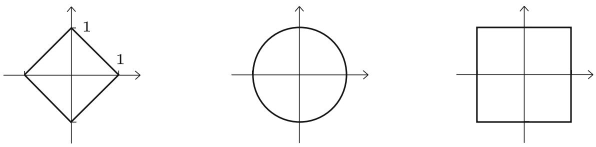
Moreover, since

$$|\|x_n\| - \|x\|| \leq \|x_n - x\| \leq c_2 \|x_n - x\|_\infty ,$$

the real-valued function $f(x) = \|x\|$ is continuous on $(X, \|\cdot\|_\infty)$, and thus on the compact set $S := \{x : \|x\|_\infty = 1\}$ it attains its minimum c_1, which is strictly positive since the norm is definite. Consequently, for all $x \neq 0$ in X, $c_1 \leq \| \|x\|_\infty^{-1} x\|$ and therefore $c_1 \|x\|_\infty \leq \|x\|$. □

In infinite-dimensional spaces the assertion of the preceding proposition does not hold. For example, let us consider on $C([0, 1])$, besides the supremum norm, the integral norm $\|f\|_1 = \int_0^1 |f(x)|\, dx$; there are sequences $f_1, f_2, \ldots$ with $\|f_n\|_\infty = 1$, but $\|f_n\|_1 \to 0$. (See Exercise 13.2.) Different norms thus yield different convergence statements.

The following pictures exhibit the unit balls of the p-norms in $\mathbb{R}^2$ for $p = 1, 2, \infty$.



$p = 1$ $p = 2$ $p = \infty$

Continuous Linear Mappings

Between finite-dimensional spaces, all linear mappings are continuous. In infinite dimensions this is no longer true. A counterexample is given by any linear mapping $T : \ell^2 \to \mathbb{R}$ satisfying $Te_n = n$ for the standard unit vectors e_n. Indeed, setting $x_n = n^{-1} e_n$ we obtain $x_n \to 0$ because of $\|x_n\| = n^{-1}$, but $Tx_n = 1 \neq 0 = T(0)$.

The continuity of a linear mapping T between normed spaces X and Y can be characterized by several equivalent properties. A number $C > 0$ is called a *bound* for a subset M of X, if $\|x\| \leq C$ for all $x \in M$; if such a bound exists, M is called *bounded* (in X). One immediately realizes that (finite) sums and scalar multiples of bounded sets are again bounded. The mapping T is called bounded on M if the image $T(M)$ is bounded in Y.

Proposition 13.2. *For any linear mapping* T *between normed spaces* X *and* Y *the following statements are equivalent:*

(i) T *is continuous on* X.
(ii) T *is continuous in 0.*
(iii) *There exists a ball* B_r *on which* T *is bounded.*
(iv) *The image* T(M) *of every bounded set* M *is bounded.*
(v) *There exists a* $C > 0$ *such that* $\|Tx\|_Y \le C\|x\|_X$ *for all* $x \in X$.

Proof. It is obvious that (i) implies (ii). To prove (ii) $\Rightarrow$ (iii) we use contraposition. Let $x_1, x_2, \ldots$ be a sequence satisfying $0 < \|Tx_n\| \to \infty$ and w.l.o.g. $\|x_n\| = r$; setting $z_n = \|Tx_n\|^{-1}x_n$ we get that $z_n \to 0$ as well as $\|Tz_n\| = 1$, thus T is not continuous in 0. To deduce (iv) from (iii), let M be bounded. Then $M \subset tB_r$ for a suitable $t > 0$ and $T(M) \subset tT(B_r)$, thus T(M) is bounded. To deduce (v) from (iv) we note that for any $x \neq 0$ we have $\|Tx\|_Y = \|x\|_X\|T(\|x\|_X^{-1}x)\|_Y \le C\|x\|_X$, if C is a bound for $T(B_1)$ in Y. To deduce (i) from (v), let $x_1, x_2, \ldots$ be a sequence with $x_n \to x$, then we have $\|Tx_n - Tx\|_Y = \|T(x_n - x)\|_Y \le C\|x_n - x\|_X \to 0$. □

Definition

By $\mathcal{L}(X; Y)$ we denote the set of all continuous linear mappings between normed spaces X and Y. When Y is the scalar field, we call it the *dual space* of X, denoted by X'. Elements of X' are called *functionals*, elements of $\mathcal{L}(X; Y)$ are called *operators*.

For any $T \in \mathcal{L}(X; Y)$, the set $\{x \in X : Tx = y\}$, with $y \in Y$ given, is a closed affine subspace of X. If, in particular, $\ell : \mathbb{R}^d \to \mathbb{R}$ is linear ($d < \infty$) and c is a scalar, we obtain hyperplanes $H = \{\ell = c\}$, decomposing $\mathbb{R}^d$ into two open half-spaces $\{\ell > c\}$ and $\{\ell < c\}$. This fact remains valid for functionals $\ell \in X'$ on arbitrary normed spaces X and serves as a starting point for geometric considerations in Banach spaces.

Since sums and scalar multiples of continuous linear mappings are again continuous and linear, X', and more generally $\mathcal{L}(X; Y)$ are vector spaces. The characterization (v) of their continuity in the preceding proposition yields that

$$\|T\| := \sup_{\|x\| \le 1} \|Tx\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\| \neq 0} \frac{\|Tx\|}{\|x\|}$$

is a finite nonnegative number; it is called the *operator norm* of $T \in \mathcal{L}(X; Y)$. We obviously have

$$\|Tx\| \le \|T\|\|x\|$$

for all $x \in X$, and $\|T\|$ is the smallest constant C with the property that $\|Tx\| \le C\|x\|$ for all x. This implies that the composition $S \circ T$ of two continuous linear mappings satisfies, because $\|(S \circ T)x\| \le \|S\|\|Tx\| \le \|S\|\|T\|\|x\|$,

$$\|S \circ T\| \le \|S\|\|T\| .$$

Proposition 13.3. *Equipped with the operator norm, $\mathcal{L}(X;Y)$ becomes a normed space. If Y is complete, $\mathcal{L}(X;Y)$ is a Banach space. In particular, the dual space X' is a Banach space.*

Proof. Definiteness holds, since $\|T\| = 0$ if and only if $Tx = 0$ for all x, which is the same as $T = 0$. Positive homogeneity and the triangle inequality are consequences of elementary properties of the supremum. Let $T_1, T_2, \ldots$ be a Cauchy sequence in $\mathcal{L}(X;Y)$. Since $\|T_n x - T_m x\| \le \|T_n - T_m\|\|x\|$, $T_1 x, T_2 x, \ldots$ is a Cauchy sequence in Y for every fixed x. When Y is complete, there exists $\lim_n T_n x =: Tx$, and one can verity (Exercise 13.4), that the mapping $T : X \to Y$ thus defined is linear and continuous, and that $T_n \to T$ in $\mathcal{L}(X;Y)$. □

On X', the operator norm is called *dual norm*, and for $\ell \in X'$ one usually terms

$$\|\ell\| = \sup_{\|x\|\le 1} |\ell(x)| = \sup_{\|x\|=1} |\ell(x)| = \sup_{\|x\|\ne 0} \frac{|\ell(x)|}{\|x\|}$$

simply *the norm of* ℓ.

Example

The formula $\ell(f) = \int f\,d\mu$ defines on $X = L_1(S;\mu)$, μ being a measure, a functional $\ell \in X'$ which satisfies $|\ell(f)| \le \|f\|_1$ as well as $\ell(1_A) = \int 1_A\,d\mu = \|1_A\|_1$ for measurable A with $\mu(A) < \infty$, thus $\|\ell\| = 1$. If moreover S is a compact metric space, μ finite, and $X = (C(S), \|\cdot\|_\infty)$, then again $\ell \in X'$, but this time $\|\ell\| = \mu(S)$, since $|\ell(f)| \le \mu(S)\|f\|_\infty$ and $\ell(1) = \mu(S)$. In particular, the Dirac measure δ_x for $x \in S$ defines a functional $\delta_x \in C(S)'$ with $\|\delta_x\| = 1$, one has $\delta_x(f) = f(x)$. (One also terms it *Dirac functional*.) On the other hand, on $X = L_1(S;\lambda)$, $S = (a,b)$, one cannot obtain a continuous linear functional from the Dirac measure δ_x, compare Exercise 13.5.

Example

If U is a closed subspace of a Hilbert space X, the orthogonal projection P_U considered in the preceding chapter defines an operator in $\mathcal{L}(X) := \mathcal{L}(X;X)$ with $\|P_U\| = 1$ whenever $U \ne \{0\}$.

Example

For a finite measure μ we consider the spaces $X = L_p(S;\mu)$ and $Y = L_r(S;\mu)$, $1 \le r < p < \infty$. If $f \in L_p(S;\mu)$, using Hölder's inequality with the decomposition $1 = r/p + (p-r)/p$ we see that

$$\|f\|_r = \left(\int |f|^r \, d\mu\right)^{\frac{1}{r}} \le \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} \left(\int 1 \, d\mu\right)^{\frac{p-r}{pr}} = C\|f\|_p, \quad C = \mu(S)^{\frac{p-r}{pr}}.$$

Therefore $L_p(S;\mu) \subset L_r(S;\mu)$ holds, and the *embedding* of $L_p(S;\mu)$ into $L_r(S;\mu)$ defined by $T(f) = f$ is linear and continuous. The inclusion is proper in general, as for example in the case $S = (0,1)$ and $\mu = \lambda$ the function defined by $f(t) = t^{-1/p}$ demonstrates.

Example

We consider an *integral operator* of the form

$$(Tf)(x) = \int k(x,y)f(y)\,\nu(dy). \tag{$*$}$$

For a given kernel k it maps a function f to a function Tf. We consider measure spaces $(S', \mathcal{A}', \mu)$ and and $(S'', \mathcal{A}'', \nu)$ as in Chap. 8 and assume that $k : S' \times S'' \to \overline{\mathbb{R}}$ is measurable. Let moreover

$$C_k := \sup_{y \in S''} \int |k(x,y)|\,\mu(dx) < \infty.$$

For any $f \in L^1(S'';\nu)$ we have

$$\iint |k(x,y)f(y)|\mu(dx)\nu(dy) \le \int C_k|f(y)|\,\nu(dy) = C_k\|f\|_1 < \infty. \tag{$**$}$$

As explained in Chap. 8, the right side of $(*)$ defines an element of $L_1(S';\mu)$. Thus, $(*)$ defines an operator $T : L_1(S'';\nu) \to L_1(S';\mu)$. T is obviously linear; moreover, it is continuous by virtue of the inequality $\|Tf\|_1 \le C_k\|f\|_1$ which is valid due to $(**)$.

Depending upon the properties of the kernel function k, integral operators of the form $(*)$ act on various different function spaces. The classical starting point is given by the Hilbert space case $T : L_2(0,1) \to L_2(0,1)$ with $\mu = \nu = \lambda$, in this case it suffices for T being continuous that $\iint |k(x,y)|^2 \, dx\, dy$ is finite.

The Dual Space of $L_p(S;\mu)$

For a given measure space $(S, \mathcal{A}, \mu)$ we consider the spaces $L_p(S;\mu)$ where $p \in [1,\infty]$. Let q be the exponent dual to p, that is, $1/p + 1/q = 1$ (here ∞ is dual to 1, and 1 is dual to ∞). When $g \in L_q(S;\mu)$, the mapping

$$f \mapsto \int fg \, d\mu$$

defines a continuous linear functional on $L_p(S;\mu)$ because

$$\Big|\int_S fg\,d\mu\Big| \le \|g\|_q\|f\|_p$$

due to Hölder's inequality. It turns out that for $p < \infty$ every continuous linear functional on $L_p(S;\mu)$ can be represented in this way. We restrict ourselves to the case where the measure μ is finite.

Proposition 13.4. *Let μ be a finite measure on a measurable space $(S,\mathcal{A})$, let $1 \le p < \infty$. Every continuous linear functional ℓ on $L_p(S;\mu)$ has the form*

$$\ell(f) = \int fg\,d\mu$$

for some $g \in L_q(S;\mu)$. The mapping $g \mapsto \ell$ is linear and isometric, that is, $\|\ell\| = \|g\|_q$ holds for the dual norm of ℓ.

In other words: The dual space of $L_p(S;\mu)$ is isometrically isomorphic to the space $L_q(S;\mu)$.

Proof. For any given $g \in L_q(S;\mu)$ we set $G(f) := \int fg\,d\mu$. As we already have seen above, G is well-defined, continuous and linear, and satisfies $\|G\| \le \|g\|_q$. The mapping $g \mapsto G$ is obviously linear. To prove the reverse inequality $\|G\| \ge \|g\|_q$ in the case $p > 1$, we consider the function

$$f = (\operatorname{sign} g)|g|^{q-1}\,.$$

We have $fg = |g|^q = |f|^p$ due to $p(q-1) = q$, and

$$\Big|\int_S fg\,d\mu\Big| = \Big(\int |g|^q\,d\mu\Big)^{1/q}\Big(\int |g|^q\,d\mu\Big)^{1/p} = \|g\|_q\|f\|_p\,,$$

so altogether $\|G\| = \|g\|_q$ in the case $p > 1$. In the case $p = 1$ we set

$$A_n = \{|g| \ge \|g\|_\infty - \frac{1}{n}\}\,, \quad f_n = 1_{A_n}\operatorname{sign} g\,.$$

We have $\|f_n\|_1 = \mu(A_n)$ and

$$\ell(f_n) = \int f_n g d\mu = \int 1_{A_n}|g|\,d\mu \ge \mu(A_n)\Big(\|g\|_\infty - \frac{1}{n}\Big) = \|f_n\|_1\Big(\|g\|_\infty - \frac{1}{n}\Big)\,.$$

It follows that $\|G\| \ge \|g\|_\infty - 1/n$ and thus $\|G\| \ge \|g\|_\infty$.

It remains to show, and this constitutes the main part of the proof, that every functional $\ell \in L_p(S;\mu)'$ can be represented in this way.

1. We want to prove that

$$\nu(A) = \ell(1_A)\,, \quad A \subset S \text{ measurable,}$$

defines a signed finite measure on $\mathcal{A}$. First, we have $\nu(\emptyset) = \ell(0) = 0$. Next, let $A_1, A_2, \ldots$ be any sequence of disjoint measurable sets. Setting $A = \cup_{n\geq 1} A_n$ we get

$$\Big\| 1_A - \sum_{n=1}^{m} 1_{A_n} \Big\|_p^p = \mu\Big(A \setminus \bigcup_{n=1}^{m} A_n\Big) \to 0$$

in the limit $m \to \infty$ due to continuity of measures, and thus, since ℓ is continuous,

$$\nu(A) = \ell(1_A) = \lim_{m\to\infty} \ell\Big(\sum_{n=1}^{m} 1_{A_n}\Big) = \lim_{m\to\infty} \sum_{n=1}^{m} \ell(1_{A_n}) = \sum_{n\geq 1} \nu(A_n)\,.$$

The set function ν therefore is σ-additive and thus a signed measure satisfying $|\nu(S)| = |\ell(1)| < \infty$.
2. Let $\nu = \nu^+ - \nu^-$ be the Jordan decomposition of ν into the measures ν^+ and ν^- according to Proposition 9.9, which are both finite since ν is finite. We have $\nu^+ \ll \mu$, $\nu^- \ll \mu$, since it follows from $\mu(A) = 0$ that $0 = \ell(1_{A'}) = \nu(A')$ for all $A' \subset A$ and therefore $\nu^+(A) = \nu^-(A) = 0$. By the Radon-Nikodym Theorem there exist densities $d\nu^+ = g^+ d\mu$, $\nu^- = g^- d\mu$, which are integrable since $\nu^\pm$ is finite. We set $g = g^+ - g^-$ and obtain for any measurable A

$$\ell(1_A) = \nu(A) = \int_A g\, d\mu$$

for a suitable integrable function g.
3. We prove that

$$\ell(f) = \int fg\, d\mu \tag{$*$}$$

for bounded measurable functions f. Indeed, $(*)$ holds for $f = 1_A$ and therefore, due to linearity, for signed elementary functions. Since the latter are dense in $L_\infty(S;\mu)$ (Exercise 13.6), $(*)$ holds as claimed.

4. We show that $g \in L_q(S;\mu)$. In the case $p > 1$ we consider the sequence of bounded measurable functions defined by

$$f_n = 1_{A_n}(\operatorname{sign} g)|g|^{q-1}, \quad A_n = \{|g| \le n\}.$$

As shown above during the proof, we have $|f_n|^p = 1_{A_n}|g|^q$ and by virtue of 3.

$$\int 1_{A_n}|g|^q \, d\mu = \int f_n g \, d\mu = \ell(f_n) \le \|\ell\| \|f_n\|_p = \|\ell\| \left(\int 1_{A_n}|g|^q \, d\mu \right)^{1/p}.$$

It follows that $\|1_{A_n} g\|_q \le \|\ell\|$ and moreover, due to monotone convergence, $\|g\|_q \le \|\ell\|$, as $|g|^q = \sup_n 1_{A_n}|g|^q$ almost everywhere. In the case $p = 1$ we set $A = \{|g| > \|\ell\|\}$ and obtain, letting $f = 1_A \operatorname{sign} g$,

$$\int 1_A |g| \, d\mu = \int fg \, d\mu = \ell(f) \le \|\ell\| \|f\|_1 = \|\ell\| \mu(A).$$

If $\mu(A) > 0$, we would have $\mu(A)\|\ell\| < \int 1_A |g| \, d\mu$ by definition of A, a contradiction. Consequently, $|g| \le \|\ell\|$ almost everywhere, thus $\|g\|_\infty \le \|\ell\|$ in the case $p = 1$.

5. Both sides of (∗) define continuous functionals on $L_p(S;\mu)$ which conincide on the dense subset L_∞ of L_p, and therefore on all of L_p. Thus we have proved the representation of ℓ as claimed.

□

The Banach Space $\mathcal{M}(S)$ of Signed Finite Measures

Let $(S, \mathcal{A})$ be a measurable space. The set

$$\mathcal{M}(S) = \{\mu \mid \mu : \mathcal{A} \to \mathbb{R} \text{ is a signed finite measure}\}$$

becomes a real vector space when equipped with the addition and scalar multiplication

$$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A), \quad (\alpha\mu)(A) = \alpha\mu(A).$$

We consider the Jordan decomposition $\mu = \mu^+ - \mu^-$ of μ into finite measures $\mu^\pm$ according to Proposition 9.9:

$$\mu^+(A) = \sup_{A' \subset A} \mu(A'), \quad \mu^-(A) = -\inf_{A' \subset A} \mu(A') = (-\mu)^+(A)$$

for measurable A. From this we immediately obtain that

$$(\mu_1 + \mu_2)^+(A) \le \mu_1^+(A) + \mu_2^+(A)\,, \quad (\mu_1 + \mu_2)^-(A) \le \mu_1^-(A) + \mu_2^-(A) \qquad (*)$$

for any $\mu_1, \mu_2 \in \mathcal{M}(S)$. The formula $|\mu| = \mu^+ + \mu^-$ defines another finite measure, called the *variation* of μ. The triangle inequality for the positive and negative part extends to the variation, because by $(*)$,

$$|\mu_1 + \mu_2|(A) \le |\mu_1|(A) + |\mu_2|(A)\,.$$

For scalar multiples we obtain $|\alpha\mu|(A) = |\alpha||\mu|(A)$ from the Jordan decomposition $\alpha\mu = (\alpha\mu)^+ - (\alpha\mu)^-$, where in the case $\alpha < 0$ we only have to take into account that $(\alpha\mu)^+ = -\alpha\mu^-$ and $(\alpha\mu)^- = -\alpha\mu^+$. It follows from the exposition above that

$$\|\mu\| = |\mu|(S)$$

defines a norm on $\mathcal{M}(S)$, since $\|\mu\| = 0$ implies that $\mu^+(S) = \mu^-(S) = 0$ and therefore $\mu = 0$. For any $\mu \in \mathcal{M}(S)$ and any measurable A we thus obtain

$$|\mu(A)| \le |\mu|(A) \le \|\mu\|\,. \qquad (**)$$

Proposition 13.5. *The space $\mathcal{M}(S)$ is a Banach space when equipped with the norm $\|\mu\| = |\mu|(S)$.*

Proof. Only the completeness remains to be proved. Let (μ_n) be a Cauchy sequence in $\mathcal{M}(S)$. For any measurable A, $(\mu_n(A))$ is a Cauchy sequence in $\mathbb{R}$ because of $(**)$. We set

$$\mu(A) = \lim_{n\to\infty} \mu_n(A)\,.$$

We want to prove that the set function μ is a signed finite measure. We have $\mu(\emptyset) = 0$. Since we may interchange the limit with finite sums, μ is finitely additive. Moreover, again because of $(**)$, one has that

$$|\mu(A) - \mu_n(A)| = \lim_{m\to\infty} |\mu_m(A) - \mu_n(A)| \le \limsup_{m\to\infty} \|\mu_m - \mu_n\|$$

for all A. In order to prove that μ is σ-additive, we consider a sequence $A_1, A_2, \ldots$ of disjoint measurable sets and set $A = \bigcup_{k\geq 1} A_k$. For all natural numbers n, l it holds that

$$\left|\mu(A) - \sum_{k=1}^{l} \mu(A_k)\right| \leq |\mu(A) - \mu_n(A)| + \left|\mu_n(A) - \sum_{k=1}^{l} \mu_n(A_k)\right|$$
$$+ \left|\mu_n\Big(\bigcup_{k=1}^{l} A_k\Big) - \mu\Big(\bigcup_{k=1}^{l} A_k\Big)\right| ,$$

where we have made use of the finite additivity of μ, already proved above. Passing to the limit superior in l while keeping n fixed yields, since μ_n is σ-additive,

$$\limsup_{l\to\infty} \left|\mu(A) - \sum_{k=1}^{l} \mu(A_k)\right| \leq 2 \limsup_{m\to\infty} \|\mu_m - \mu_n\| .$$

Passing once more to the limit superior, this time in n, gives 0 on the right side, and consequently $\mu(A) = \sum_{k\geq 1} \mu(A_k)$. □

The Dual Space of $C(S)$

Let S be a compact metric space, equipped with the Borel σ-algebra $\mathcal{B}$, and let $C(S)$ be the Banach space of all real-valued continuous functions on S. By the representation Proposition 11.3, we may represent every positive linear functional ℓ on $C(S)$ as an integral with respect to some finite measure μ. If we also allow signed measures, we can find such a representation for arbitrary continuous linear functionals on $C(S)$.

A signed finite measure μ is called *regular*, if μ^+ and μ^- are regular (or equivalently, if $|\mu|$ is regular). From Proposition 7.6 it follows that every signed finite measure on the compact metric space S is regular.

Proposition 13.6. *Let S be a compact metric space. Every continuous linear functional ℓ on $C(S)$ can be uniquely represented in the form*

$$\ell(f) = \int f \, d\mu$$

with a signed finite regular measure μ. The mapping $\mu \mapsto \ell$ is linear and isometric, that is, $\|\ell\| = \|\mu\|_{\mathcal{M}(S)}$ holds for the dual norm of ℓ.

Proof. For any given $\mu \in \mathcal{M}(S)$ the functional $\ell : C(S) \to \mathbb{R}$, given by

$$\ell(f) = \int f \, d\mu \, ,$$

is linear. Due to the Jordan decomposition $\mu = \mu^+ - \mu^-$ we obtain the estimate

$$|\ell(f)| \le \int |f| \, d\mu^+ + \int |f| d\mu^- \le \|f\|_\infty \|\mu^+\| + \|f\|_\infty \|\mu^-\| \le \|f\|_\infty \|\mu\| \, ,$$

and thus ℓ is continuous with $\|\ell\| \le \|\mu\|$, therefore $\ell \in C(S)'$. In order to prove the reverse inequality $\|\ell\| \ge \|\mu\|$, let A_+ and $A_- := A_+^c$ be the sets belonging to the Jordan (resp. Hahn) decomposition satisfying $\mu^+(A_-) = \mu^-(A_+) = 0$. Since μ is regular, for arbitrary $\varepsilon > 0$ we find compact sets $K_+ \subset A_+$ and $K_- \subset A_-$ such that $\mu^\pm(A_\pm) \le \mu^\pm(K_\pm) + \varepsilon$. We now define the continuous functions

$$f_\pm(x) = (1 - \alpha^{-1} d(x, K_\pm))^+ \, , \quad f = f_+ - f_- \, ,$$

where $\alpha := \operatorname{dist}(K_+, K_-) = \inf_{x_\pm \in K_\pm} d(x_+, x_-)$. We have $f = 1$ on K_+, $f = -1$ on K_-, and $\|f\|_\infty \le 1$. We estimate

$$\begin{aligned}
\int_S f \, d\mu &= \int_{K_+} f \, d\mu + \int_{K_-} f \, d\mu + \int_{(K_+ \cup K_-)^c} f \, d\mu \\
&\ge |\mu|(K_+) + |\mu|(K_-) - |\mu|((K_+ \cup K_-)^c) = 2(|\mu|(K_+) + |\mu|(K_-)) - |\mu|(S) \\
&\ge 2(|\mu|(A_+) + |\mu|(A_-) - 2\varepsilon) - |\mu|(S) = |\mu|(S) - 4\varepsilon = \|\mu\| - 4\varepsilon \, .
\end{aligned}$$

Therefore, $\|\ell\| \ge \ell(f) \ge \|\mu\| - 4\varepsilon$, and consequently $\|\ell\| \ge \|\mu\|$, letting $\varepsilon \to 0$. The isometry $\|\ell\| = \|\mu\|$ just proved implies the uniqueness of μ in the representation of ℓ, since the mapping $\mu \mapsto \ell$ is obvious linear.

It remains to show that such a μ exists for any given $\ell \in C(S)'$. In order to achieve this, we represent ℓ as the difference of two positive linear functionals and apply the Riesz representation Proposition 11.3. We define

$$\ell^+(f) = \sup_{0 \le \varphi \le f} \ell(\varphi) \, , \quad \text{if } f \ge 0.$$

For any such φ we have $\|\varphi\|_\infty \le \|f\|_\infty$, therefore $\ell(\varphi) \le \|\ell\| \|\varphi\|_\infty \le \|\ell\| \|f\|_\infty$, and thus $0 \le \ell^+(f) < \infty$ for $f \ge 0$. Immediately from the definition we obtain that

$$\ell^+(f) + \ell^+(g) \le \ell^+(f + g) \, , \quad \ell^+(\alpha f) = \alpha \ell^+(f) \, ,$$

for any $f, g \geq 0$ and any $\alpha \geq 0$. In order to prove the reverse inequality, let $0 \leq \varphi \leq f + g$. We have

$$\varphi = \min(\varphi, f) + (\varphi - f)^+ , \quad (\varphi - f)^+ \leq g ,$$

and therefore

$$\ell(\varphi) = \ell\big(\min(\varphi, f)\big) + \ell((\varphi - f)^+) \leq \ell^+(f) + \ell^+(g)$$

by the definition of ℓ^+. Passing to the supremum with respect to φ yields the inequality $\ell^+(f + g) \leq \ell^+(f) + \ell^+(g)$. We conclude that

$$\ell^+(f) + \ell^+(g) = \ell^+(f + g) , \quad \text{if } f, g \geq 0. \tag{$*$}$$

We now define, for arbitrary $f \in C(S)$,

$$\ell^+(f) = \ell^+(f^+) - \ell^+(f^-) .$$

The linearity of ℓ^+ on $C(S)$ is proved in the same manner as for the Lebesgue integral, namely we apply ℓ^+ in view of $(*)$ to the identities

$$(f + g)^+ + f^- + g^- = (f + g)^+ + f^+ + g^+ ,$$
$$(-f)^+ + f^+ = (-f)^- + f^- .$$

Besides ℓ^+, also $\ell^- := \ell^+ - \ell$ is a positive linear functional on $C(S)$. The Riesz representation Proposition 11.3 yields finite measures μ_+ and μ_- such that

$$\ell^+(f) = \int f \, d\mu_+ , \quad \ell^-(f) = \int f \, d\mu_- .$$

Finally, $\mu = \mu_+ - \mu_-$ yields the sought-after representation of ℓ. □

Exercises

13.1 Prove that the space $C(S)$ of continuous functions on a compact metric space S, equipped with the supremum norm $\|f\|_\infty = \sup_{x \in X} |f(x)|$, is a Banach space.

13.2 Let M be a subset of a normed space X. Prove that the distance function $d(x, M) = \inf_{z \in M} \|x - z\|$ is nonexpansive when viewed as a function of x.

13.3
1. Let $T : X \to Y$ be a linear mapping between normed spaces spaces X and Y. Prove that if X is finite-dimensional, then T is continuous.
2. Prove that every finite-dimensional normed space is a Banach space.

13.4 **Completeness of $\mathcal{L}(X; Y)$**

Let X, Y be Banach spaces, let $T_1, T_2, \ldots$ be a Cauchy sequence in $\mathcal{L}(X; Y)$, let $T : X \to Y$ be defined by $Tx = \lim_{n \to \infty} T_n x$. Prove:

(i) T is linear.

(ii) The set $\{\|T_n\|\}_{n\in\mathbb{N}}$ is bounded, and T is continuous.
(iii) $\lim_{n\to\infty} \|T_n - T\| = 0$.

13.5 We interpret the set of continuous functions $f : [0, 1] \to \mathbb{R}$ as a subspace U of the Banach space $L_1([0, 1]; \lambda)$, equipped with the L_1 norm. Prove:
(i) U is not closed in X, and therefore not complete.
(ii) Let $x \in [0, 1]$. The functional defined by $\delta_x(f) := f(x)$ is not continuous on U.

13.6 Let $(S, \mathcal{A}, \mu)$ be a measure space. Prove that for every $f \in L_\infty(S; \mu)$ there exists a sequence $f_1, f_2, \ldots$ of signed elementary functions such that $\|f_n - f\|_\infty \to 0$.

Bibliography

1. H. Bauer, *Measure and Integration Theory* (de Gruyter, Berlin/New York, 2001)
2. J. Elstrodt, *Maß- und Integrationstheorie*, 6. Aufl. (Springer, 2009)
3. L.C. Evans, R.F. Gariepy, *Measure Theory and Fine Properties of Functions* (CRC, Boca Raton, 1992)
4. P.R. Halmos, *Measure Theory* (Van Nostrand, New York, 1950/Springer, New York, 1974)
5. A. Klenke, *Probability Theory*, 2nd edn. (Springer, London/New York, 2013)
6. A. Pietsch, *History of Banach Spaces and Linear Operators* (Springer, London, 2007)
7. W. Rudin, *Principles of Mathematical Analysis*, 3rd edn. (McGraw-Hill, New York, 1976)
8. W. Rudin, *Real and Complex Analysis*, 3rd edn. (McGraw-Hill, Singapore, 1986)
9. R. Schilling, *Measures, Integrals and Martingales* (Cambridge University Press, Cambridge/New York, 2005)
10. The MacTutor history of mathematics archive, http://www-history.mcs.st-and.ac.uk/
11. D. Werner, *Funktionalanalysis*, 6. Aufl. (Springer, 2007)
12. D. Werner, *Einführung in die Höhere Analysis*, 2. Aufl. (Springer, 2009)

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks in Mathematics, DOI 10.1007/978-3-319-15365-0

Index

M. Brokate, G. Kersting, *Measure and Integral*, Compact Textbooks
in Mathematics, DOI 10.1007/978-3-319-15365-0